

# A New Approach to Research Data Management with a Focus on Traceability: Adacta

Hendrik Gossler<sup>1,\*</sup>, Johannes Riedel<sup>1,2</sup>, Eric Daymo<sup>3</sup>, Rinu Chacko<sup>2</sup>, Sofia Angeli<sup>2</sup>, and Olaf Deutschmann<sup>2,\*</sup>

DOI: 10.1002/cite.202200064



This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



Supporting Information  
available online

Traceability between samples, devices and data is of great importance to the catalysis community. Adacta is a new research data management (RDM) system designed to create a traceable digital twin of a testing environment, not only storing data, but also creating a readily retrievable time-accurate record of the critical components used to measure catalyst performance. Future developments include extending Adacta to interact with electronic laboratory notebooks and to spawn simulations directly using data and measurements stored in the database.

**Keywords:** Digital catalysis, FAIR principles, Research data management, Software, Traceability

*Received:* May 16, 2022; *revised:* July 22, 2022; *accepted:* August 23, 2022

## 1 Introduction

Industry 4.0 is poised to integrate real time data, machine learning, automation, and communication in a way that generally improves the productivity of manufacturing and other commercial operations. Already, inexpensive cloud data storage and the proliferation of sensors has led to the realization of digital twins and the Industrial Internet of Things [1]. Research and development will also benefit from this new era of ubiquitous data acquisition and storage. Research data management (RDM) is already required for most academic and industrial projects to ensure data are available for future use. General guidelines to assist researchers with improving the reusability of their research data were introduced by Wilkinson et al. [2] as the FAIR (Findability, Accessibility, Interoperability and Reusability) data principles. Even though traceability and reproducibility are not always discussed in detail in RDM plans, to ensure the reusability of a dataset, traceability is an essential prerequisite. As a result, a dataset should be accompanied by metadata that is as rich as possible. Additionally, it is crucial to define well-structured (meta)data and widely used controlled vocabularies and ontologies to enhance interoperability.

When good RDM practices are followed, the benefits to chemical process technology and catalyst development are promising. Schlögl [3] envisions Catalysis 4.0, wherein catalysis science and big data are integrated to find rational designs for new catalysts. However, as noted by Wulf et al. [4], catalysis is an interdisciplinary science coupling work-

flows associated with surface science and chemical process development as well as encompassing a broad range of characteristic time and length scales. Moreover, catalysis research includes manual laboratory tasks, the use of measuring devices that do not provide digital recordings, plus a wide variety of experimental reactor configurations used for the assessment of catalytic performance. Thus, RDM in a catalysis development environment is not as simple as only utilizing an electronic laboratory notebook (ELN) or a data acquisition system. This is because the development of a catalyst naturally results in a complex and rich data structure, which is difficult to capture in the form of a digital twin using a traditional ELN and/or data historian. An example of this rich data structure was described by Weber et al. [5], which focused on the characterization data and metadata (e.g., instruments, operating parameters) required to create a digital twin of a commercial CO<sub>2</sub> methanation

<sup>1</sup>Dr. Hendrik Gossler, Johannes Riedel  
h.gossler@omegadot.software  
omegadot software & consulting GmbH, Mühlweg 40, 67117 Limburgerhof, Germany.

<sup>2</sup>Johannes Riedel, Rinu Chacko, Dr. Sofia Angeli,  
Prof. Dr. Olaf Deutschmann  
deutschmann@kit.edu  
Institute for Chemical Technology and Polymer Chemistry, Karlsruhe Institute of Technology, Engesserstraße 20, 76131 Karlsruhe, Germany.

<sup>3</sup>Eric Daymo  
Tonkomo LLC, Gilbert, Arizona 85297, United States.

catalyst. Considering these complex data structures and the need for reproducibility, the best practices for handling catalysis data were described by Trunschke et al. [6], which included the recommendation to document catalyst synthesis, activation, and testing data along with metadata in a complete and standardized way.

Additionally, post processing calculations can also be the cause of errors (or error propagation) and should be carefully assessed in a comprehensive catalysis RDM system. An example of what can be done with data that are in a well-structured format is CaRMeN (catalytic reaction mechanism network) [7, 8]. CaRMeN serves as a structured archive in machine-readable format that enables long-term storage of catalysis performance data. By automating the reaction kinetics simulation workflow and being coupled with a wide range of reactor simulation tools, such as DETCHEM [9], CaRMeN can generate experiment vs. simulation comparisons at a large scale.

Recognizing the benefit of RDM for catalyst development, the UK Catalysis Hub is designing and implementing software and equipment to manage large amounts of catalysis research data at its member institutions [10, 11]. Furthermore, coordinated efforts under the NFDI4Cat [12] among the NFDI initiatives, funded by the German Research Foundation (Deutsche Forschungsgemeinschaft), are focused on the development of tools to collect, process, preserve and reuse catalysis data at German academic and industrial institutes. The NFDI4Cat project is also training researchers and helping to create everyday workflows to assist in the digitalization of catalysis information. In addition to these consortia, there are also commercial tools for catalysis development data management, such as previously mentioned ELNs and data historian and database software, like myHTE [13, 14], although in the case of myHTE the RDM software appears tightly coupled to their equipment offerings.

While the catalysis community has access to various tools to assist with data organization and management, all the necessary capabilities to create a digital twin of the catalysis development and testing processes are generally not yet integrated into a single product and therefore require different workflows and data storage requirements. Further, none of the existing product offerings appear to create digital twins of experimental systems, nor does there seem to be an emphasis on traceability between specific equipment (i.e., a device with a unique serial number) and data. Finally, not all of the RDM solutions for the catalysis community are commercial and thus widely available.

Adacta fills this gap in RDM software offerings, applying the FAIR data principles along with best practices learned from other projects (e.g., CaRMeN [7, 8]) to ensure the rich data structures associated with catalysis research and development are preserved for future use. As such, Adacta is neither data acquisition system/historian nor an ELN, and thus does not fall neatly into existing RDM software categories. Rather, the software is written with a particular focus on

traceability and consistency by creating a digital twin of the catalyst and associated testing systems. Adacta is also highly flexible in that it does not rely on rigid predefined “forms” to input accompanying metadata which, in the authors’ experience, represents a maintenance burden due to the many customized forms needed to support a wide variety of experimental setups that can evolve with time. This flexibility is the main reason Adacta is applicable for use outside of the catalysis development community. Yet, Adacta enforces data structure and organization, for example, by connecting data to specific devices. Eventually, Adacta will also serve as a launch point for simulations involving the stored data. Plus, the software is being developed as a commercial and professional product that can be licensed for both academic and industrial environments.

At this time the features needed by Adacta to create a digital twin of a catalyst test stand are finished, and the description of these capabilities will be discussed in detail in the next sections. Specifically, the workflow of how a user adds information into the Adacta database, and then interacts with the stored information to discover relationships between devices, samples and resources will be explored. Afterward, the outlook for Adacta will be reviewed, including plans for additional features that will expand the scope of the software’s digital twin capability.

## 2 Terminology and Architecture

Perhaps the greatest challenge in storing research data lies in being able to understand the data at a later point in time by others not having been involved in its creation. This challenge in part motivated the development of the previously mentioned FAIR Guiding Principles for scientific data management and stewardship [2]. The FAIR guidelines are intended to improve the “Findability, Accessibility, Interoperability and Reuse of digital assets” (Ibid.). FAIR further emphasizes the use of computers to find and utilize data with minimal human intervention, recognizing that the volume and complexity of data will only grow with time.

Part of the solution to this challenge of seamless and automated data management is consistent data quality and presentation. For example, each series of data needs metadata describing their units. However, the system must be flexible during the data import process because data can be organized by acquisition systems in many different ways. There must also be access control, such as a mechanism that gives the creator of the data the possibility to control which other users have access (including read-only or read/write permissions). Such access controls are important for many reasons, including complying with non-disclosure agreements and controlling intellectual property. Finally, the physical storage location for data needs to be addressed. Some organizations will prefer storing data in the cloud, but others will want to ensure control of their data resides on their own servers. Adacta’s design that addresses these com-

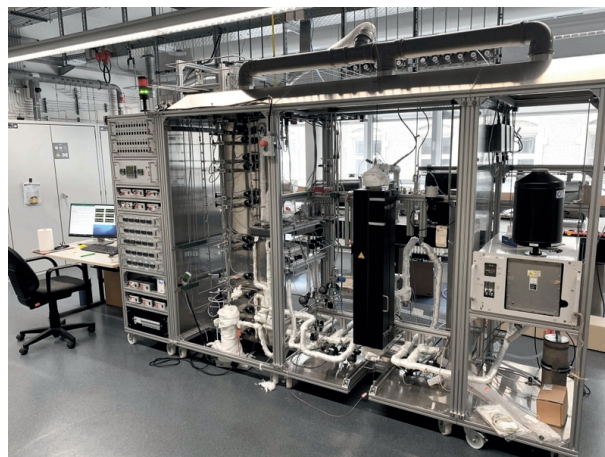
plex issues is outlined in the following sections. But first, it is important to review key terminology used within the software.

Although designed with the catalysis community in mind, Adacta is a general tool to create a digital twin of any type of experimental system. Thus, the terminology is broad in order to encompass not just laboratory-scale chemical processes, but any experimental system that utilizes *devices* (e.g., controllers, instrumentation, sound recorders, video cameras, etc.) to collect data. As such, experimental data and derived data from any post-processing steps are termed *resources*. The processes by which new data are obtained from existing data in the course of post-processing are called *transformations*. Existing data here can be either the original resource as it was imported (raw data) or the result of a previous transformation.

Transformations are critical to summarize data for use in subsequent analyses and simulations. Conversion of volumetric flow rate to velocity and time averaging data as the input to a simulation are two simple examples of transformations. Further examples include calculating conversion, yield, or selectivity from measured concentrations. A high degree of flexibility comes from the ability to arbitrarily mix and match many simple but succinct transformations. Transformation steps are typically created for an isolated elementary purpose so that they can be applied in various contexts. Multiple transformation steps can be chained by sequentially acting on the result of a previous step. However, transformations need not always be simple as in the preceding examples. Since they are implemented as code, they can be arbitrarily complex, which comes at the cost of reusability, however.

A device may be an individual piece of equipment, or it could be an ensemble of *components* that are used to collect a resource's data. Thus, a device can range from a single mass flow controller to a test rig (aka a test stand), which is typically assembled using many components (an example of which is shown in Fig. 1). In this context, a component is a child device that functions as part of the operation of a higher-level parent device, much like a mass flow controller (a device) is a component of a test stand (a higher-level device in the database structure).

Device information includes not only *specifications* related to the piece of equipment (such as the manufacturer and model number), but also item-specific information such as the unit's serial number. Devices that collect the same type of information, like mass flow controllers can be combined to define a *device type*. Device types are also associated with their units of measurement (e.g., standard liters per minute, slpm) so that in the future transformations and other user interactions can happen more seamlessly by ensuring the user is not presented with interactions that are inappropriate for a given device type. For example, no transformations or other dialogue boxes associated with temperature measurement will be provided to the user when interacting with a mass flow controller device type.



**Figure 1.** A catalyst test stand at the Exhaust Gas Center of the Karlsruhe Institute of Technology.

This linkage of components to devices, devices with resources, plus the connection of samples to devices and resources, is a fundamental aspect of Adacta's architecture. Because experimental data are generally date and time stamped, Adacta provides a historical record of which specific devices and samples were used to collect the data. Moreover, if a device is swapped out during an experimental run, Adacta can be used to ensure a permanent record is kept of the exact devices used to collect data at any point in the resource's timeline.

Meanwhile, an experimental device, like a reactor test stand, is often assembled to evaluate the performance of a *sample*. As an example, a sample might be a catalyst, and all samples are associated with a creator who prepared the sample or is otherwise responsible for its disposition. Samples can be installed in different devices if needed, and, like devices, samples can also have specifications, such as manufacturer (if applicable).

A question arises as to why abstract terms such as resource and device are used instead of more colloquial terms, such as data or equipment. Devices are intended to include a wide range of physical hardware, from a test stand that is made up of many components, down to the individual devices that comprise the test stand. For example, a test stand is a device that may contain a flow controller as a component, but that flow controller component of the test stand is also a device. Using the term device makes Adacta more universal in its treatment of physical hardware so that the software is applicable to the widest range of applications. Similarly, the term resources broadens the applicability of Adacta to any situation where information and hardware are associated, because resources are not limited to tabulated data sets, but can include media such as sounds or pictures. Likewise, the term transformation is used instead of calculation to imply that Adacta is not limited to traditional spreadsheet operations. Since resources encompass more than just tabular data, transformations are similarly

broader than simple row and column manipulations. For example, transformations may include operations on any sort of media, like filtering an image or sound clip. Ultimately, the universal nature of the application is best communicated by not using terminology specific to the scientific process of any particular discipline with a bias toward tabular data. The choice of such new terms may also free the user from preconceived notions of what more commonly used terminology may conventionally imply about what counts as data, devices, or calculations.

### 3 Traceability of Data and Workflow

Traceability refers to 1) being able to trace a set of data back to the device that was used to record it, and 2) to trace a set of data that has undergone post-processing back to the raw data initially imported into Adacta. Both points are equally important and critical to understandable data.

To ensure traceability, the general workflow in Adacta revolves around keeping an up-to-date representation of an experimental setup in the database and associating time-based events to the setup. The knowledge of a component's installed location within a test stand at any given time is what allows Adacta track the relations between components and samples. As an example, Adacta is able to "compute" that a given set of data (a resource) is related to a sample not because the user specifically specifies this connection at import time, but rather because the time of the recording of a device is known and it is also known when the sample was installed in another device related to the component hierarchy. It is therefore crucial to understand that time is the primary key that Adacta uses to establish connections in its internal database. Using time as the primary instrument to establish connections opens the door to many interesting features and has shown to be an extremely flexible approach to storing data, which gives rise to Adacta's universal and interdisciplinary applicability.

It is also important to note that Adacta associates all measured data with the device that makes the measurement, not with the sample that is being measured. This is because all measurements are made by instruments that are prone to random and systematic errors, and it is vital to be able to trace data to specific measurement devices over time. This approach may seem intuitive for online instrumentation like flowmeters, but it also applies to offline tests like measuring the mass of the sample or obtaining a FTIR spectra. So, for example, a sample's mass is not an attribute of the sample, but rather the sample's measurement is associated with the balance that makes the measurement. In the specific example of a sample's mass, first a sample is associated with a test stand (device) at a point in time. If a mass measurement is made of the sample after an experiment (e.g., to measure weight gain in an exhaust-gas aftertreatment experiment at elevated temperature), then in a second decoupled step, the user places a sample on a balance at a

certain point in time and imports the value from the scale into Adacta as a new resource. Not only does this approach mimic the actual workflow whereby the experimenter puts a sample on the balance and records the mass, but by importing data to Adacta in this way the date and time of the measurement is memorialized, along with information about the scale that was used. This method ensures full traceability of both the sample mass and the measurement device (in this case, the balance). Additionally, because the details about the balance that was used to record the mass are also captured as metadata, it is possible to trace back to all affected resources (including other data and transformations) in case a fault is later discovered in the equipment. The Adacta workflow reflects that all measurements are prone to errors or inaccuracies due to random and systematic errors and thus are usually not exact values. Thus, the best approach for ensuring traceability is to associate all measurements with the device that actually makes the measurement, and to use time to establish the relationships between resources, devices and samples.

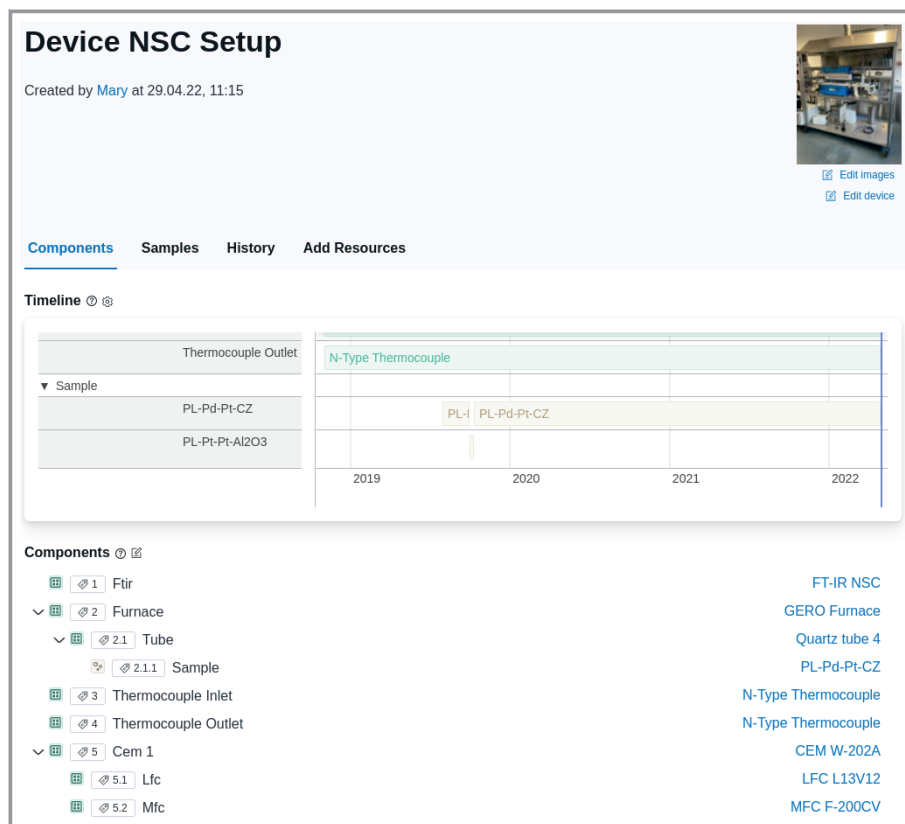
#### 3.1 Importing Resources

Given that the device hierarchy determines the relations between data and samples, the test stand must be created in Adacta before resources can be imported. For the same reason it is essential that the test stand is accurately maintained in case changes (e.g., swapping of components) are made. In a second step, the user may import data and then associate each data series within that file with the component of the stand which was used to record it.

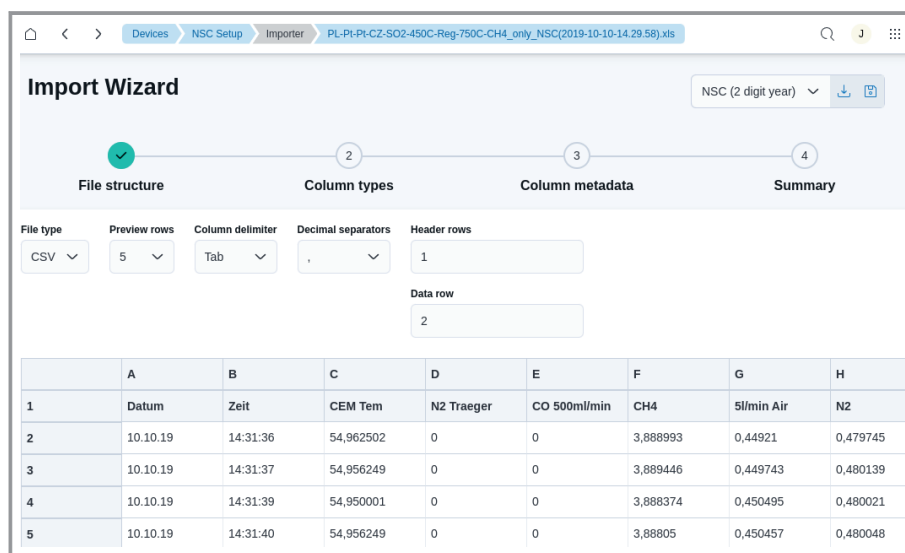
A typical resource that can be imported is a comma-separated (.csv) or tab-separated (.tsv) data table obtained from a data acquisition system, like LabView. An import wizard guides the user through a multistep import process. In general, the import process is initiated in the Device View (Fig. 2), so that the import wizard can provide the necessary context and make appropriate suggestions. These include hints on which data series can be associated to which device or optionally a picture of the assembly of components, as well as a list of all components and devices that comprise the data collection unit.

The import wizard (Fig. 3) also assists with the association of dependent variables (e.g., measurements) to independent variables (e.g., date and time stamp), as well as with the devices used to collect each column of information. This process can be tedious for a data set with many devices, but to reduce the burden on the user common patterns are recognized. For example, if a given data column is given units associated with mass per time (e.g.,  $\text{kg s}^{-1}$ ), Adacta realizes that the equipment in question is associated with some sort of flow control or measurement.

If the data structure of a resource will be reused in future experiments, a preset can be created. Then, the import of the next resource with the same tabular configuration hap-



**Figure 2.** Selecting the device “NSC” provides a full list of its components, as well as the ability to associate new resources with this device (via the “Add Resources” link).



**Figure 3.** The Adacta Import Wizard assists with the addition of new resources to the Adacta database.

pens almost instantly. Eventually Adacta’s database will be populated with many resources that can be revisited for future analysis (Fig. S1 in the Supporting Information).

device setup view, the associated component device is displayed.

### 3.2 Viewing and Editing Devices

The list of devices stored in the database can be viewed and edited via the Device View (Fig. 4). This view also allows the user to see which components are currently installed in which device (e.g., test stand). Selecting a device or component reveals a timeline (Fig. 5) that visualizes the locations where the selected apparatus has been installed over time, and what samples are associated with the equipment. If a device is made up of components, then the components which comprise the device are also shown on the timeline. At the bottom of the timeline are icons that allow the components that comprise a device (e.g., the flow controllers in a test stand) to be added, deleted, or edited (Fig. S2). For the avoidance of doubt, this is not the only way to edit a device’s or component’s information. Other contexts that also present the user with tabulated or graphical information about components and devices are also entry points to the device editing dialogue boxes.

Additional views in the device window allow the user to see which samples are associated with the device, the device specifications, and the device’s history (Fig. 6). It is noted that photos of devices are visible in Figs. 5–6 and Figs. S2–S3, and these photos can be readily changed by the user in the Device View. Finally, from within the device timeline view, an image describing the setup can be attached (Fig. 7). This user-supplied setup diagram can be further edited to place tags on the various components that comprise the device. When these components are clicked on the

Name ↑	Manufacturer	Creator	Currently installed in	Actions
Alumina tube 1		Mary		Show
Alumina tube 2		Mary		Show
Alumina tube 3		Mary		Show
Alumina tube 4		Mary		Show
CEM W-202A		Robert	NSC Setup	Show
CEM W-303B		Robert	TWC Setup	Show
FT-IR NSC		Mary	NSC Setup	Show
FT-IR TWC		Mary	TWC Setup	Show
GERO Furnace		Mary	NSC Setup	Show
HTM Reetz Furnace		Mary	TWC Setup	Show

Figure 4. The Device View lists all the devices and components in the Adacta database.

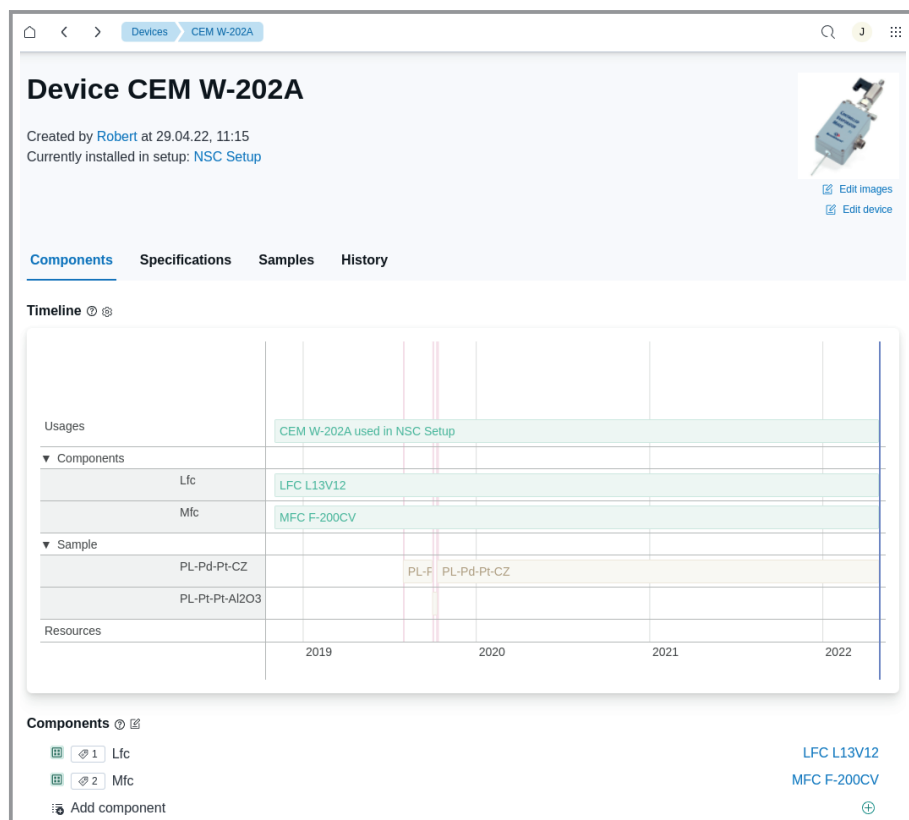


Figure 5. The device timeline view shows any associated subcomponents, and the relationship between the present device and samples. The list of components at the bottom of the Device View is a launching point to add, delete or edit a device.

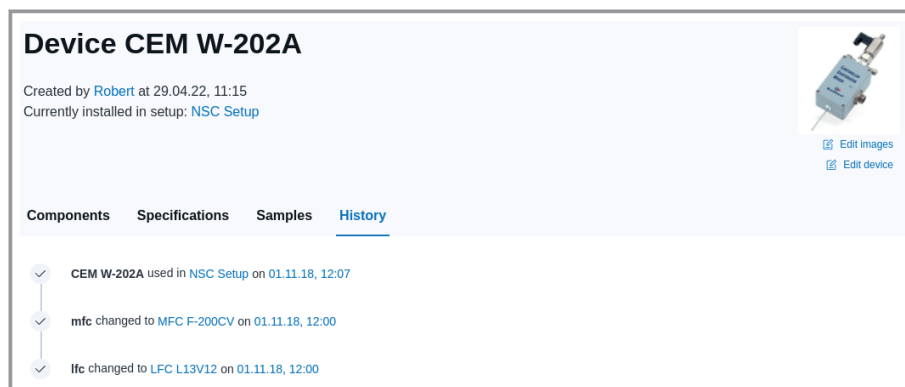
### 3.3 Viewing and Editing Samples

Similar to the Device View, the Samples View allows the user to see all of the samples in the database, including the

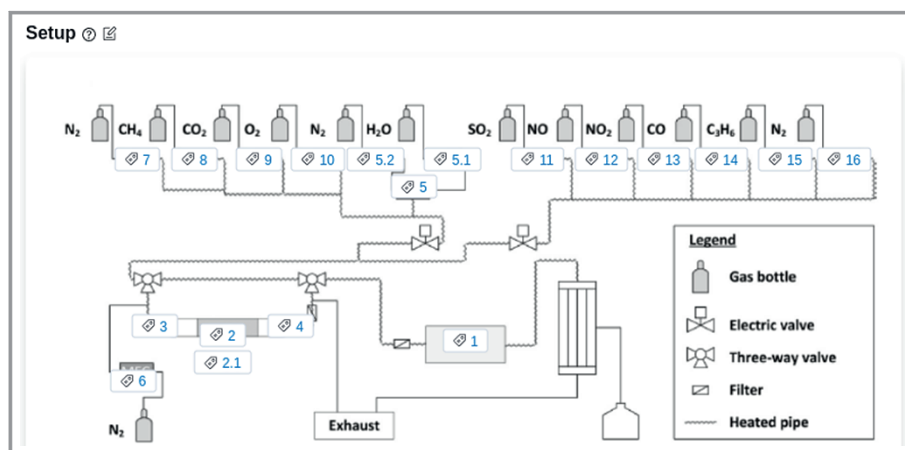
currently installed location. Samples are generally installed in devices (e.g., a catalyst sample can be installed inside a quartz tube), and so it is possible to link to the installed device from the Samples View (Fig. S3). Also like the Device View, clicking the “show” link associated with each sample brings up a timeline showing where the sample has been installed over time. Devices directly associated with a sample (such as a quartz tube that contains a sample during a test) can be accessed through the sample timeline (Fig. 8), and from there all the devices associated with this sample’s testing can be found and displayed. The Sample View timeline also contains a data link, which leads to a preview of all device data associated with the testing of this sample (Fig. 9). Clicking on the “show” link on any Device’s data preview window will display an expanded view of the collected data (Fig. 10).

### 3.4 Data Repositories

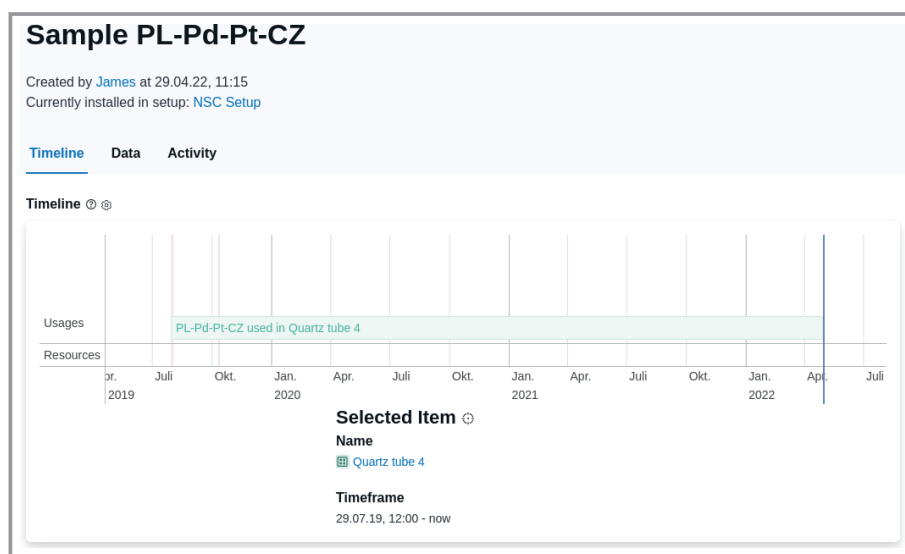
Adacta can link to one or more data repositories from the user settings tab (Fig. S4). Data repositories control the physical location where the resources and related metadata are stored. Therefore, depending on user requirements, data can be stored locally on a single computer, or remotely on a company or department server, or using Omegadot’s servers. If data are stored on remote servers, then a user can access from multiple computers (e.g., office and laboratory workstations) all the same databases that contain the user’s resources and metadata associated with devices and samples.



**Figure 6.** The Device View allows a user to easily access the history of each piece of equipment in the database.



**Figure 7.** The device flowsheet is accessible from the Device View timeline. The tags are all selectable and pull up the Device View for the selected component.



**Figure 8.** A timeline view (accessible within the Sample View) displays an overview of where samples are used over a period of time. For example, here PL-Pd-Pt-CZ is installed in Quartz Tube 4 over the displayed timeline, and if needed, the user can click on any of the blue-colored links to drill down into details on the quartz tube.

## 4 Data Traceability Capabilities and Searching

The previous discussion explained how the connections between resources, devices and samples can be interrogated. These connections make it possible to quickly determine the answers to the following inquiries:

Which resources in the database were recorded using a specific sample within a given time-frame?

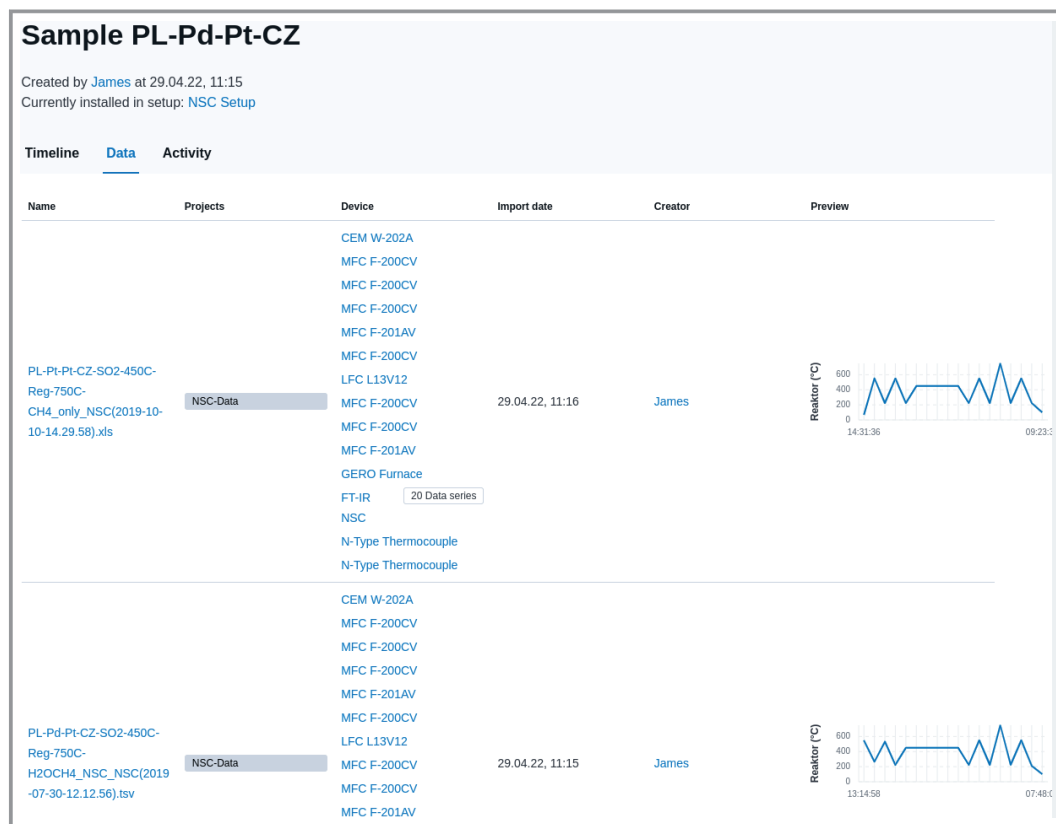
What resources were recorded at a specific time?

What measurements were done by a specific user?

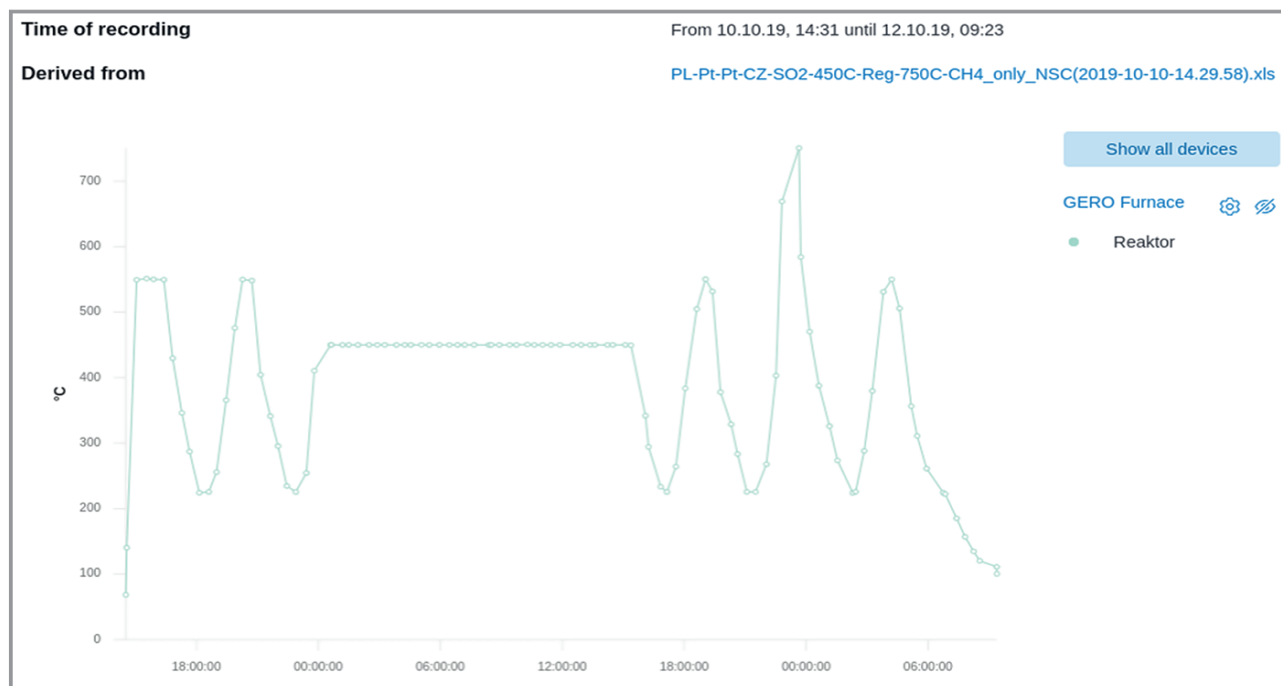
Where was a particular device used at a specific time?

As shown, answers to these questions can be found through visualization features, such as representing the temporal relationships between resources, devices and samples on a timeline. Further, if data are recorded during a single experiment using two or more independent devices, then the resulting resources can be visualized together in a joint graph. Multiple resources can also be displayed in one graph for comparison, even if the resources were gathered at different times.

The usefulness of such visualization and search features can be critical for ensuring data quality and for providing the necessary context for using/interpreting recorded data. For example, if an impurity is found in a tested sample or if there is a malfunction/miscalibration of a device, it is possible to quickly identify all affected experiments. As a result, it is now possible to readily trace the exact setup of the experimental system and determine where and how a recorded or calculated value was determined. The fact that the design and all subsequent changes are also recorded for each test stand means that it



**Figure 9.** The data link under the Sample View timeline previews the resource information (e.g., experimental data) for every device associated with the present sample. Clicking the sample name displays a larger view of the experimental data (Fig.10).



**Figure 10.** Resource data can be plotted from within the Resource View.



is also possible to trace the step-by-step development of the test stand at a later date. In combination with the link to the recorded resources, it is also possible to evaluate changes or modifications that have been made. For example, one can readily see if a change to any device or sample exhibited the expected response. The functionalities and visualizations enabled by traceability already offer a direct practical advantage to the users. Thus, immediate usefulness of Adacta creates an incentive for a researcher to use the software and ensure its database is kept up to date. This lays a foundation for establishing novel RDM software, since RDM can only realize its full potential if it is actively used and fed with data.

Currently, the integrated full-text search allows users to easily find all devices, resources, and samples from anywhere in the app. The requirements to couple Adacta to ELNs and simulation tools such as CaRMen [7, 8] can be readily met assuming exchange of information between Adacta and other electronic tools can be accomplished via text or tabulated files. Accessing metadata normally found in ELNs would allow Adacta to quickly trace suspect data to simulations and other analyses. Thus, as an example, the impact of an identified impurity in a test sample will eventually be traceable from metadata in an ELN entry through experimental testing and simulations based on data associated with the suspect sample.

## 5 Summary and Outlook

Presently, Adacta is capable of basic data management and storage, allowing users with the appropriate permissions to add resources, devices and samples, then follow the relationships and hierarchies associated with these attributes across time and resources (Fig. 11). The previous sections listed many ways in which information stored can be linked

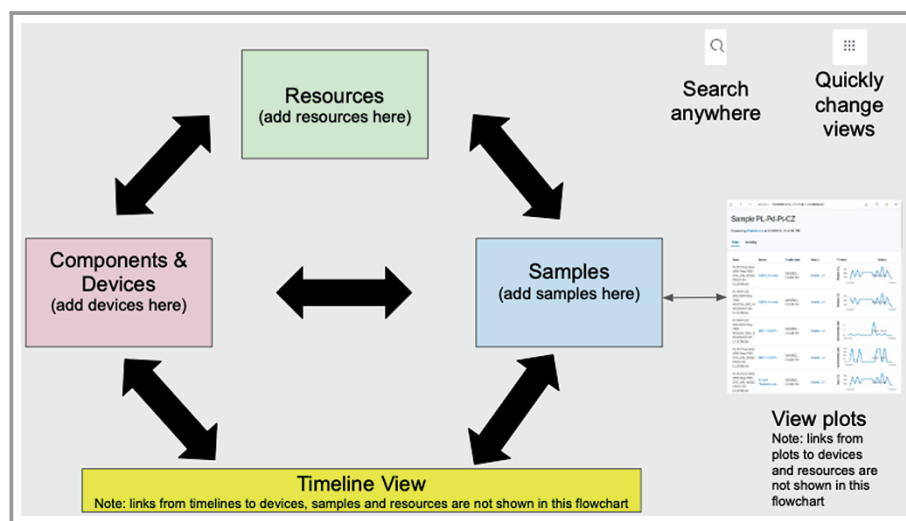
and accessed from a variety of different views, which is important because the relationships between resources, samples and devices can be complicated, and it is convenient to use links to immediately switch the focus between different information found within the database. Data can also be viewed in different physical locations, so it is possible to conduct data analysis from the comfort of one's office without needing to separately replicate files outside the Adacta system for access to results stored on the test stand computer.

There are several key developments underway to advance the concept and software: transformations, links to electronic laboratory notebooks and other resource media, coupling with simulation tools, and data security.

As previously mentioned, transformations are mathematical calculations that operate on resources or other data stored in the database. Since resources include data and transformations include calculations involving data (e.g., the conversion of volumetric flow rate at standard conditions to actual velocity), then the output of a transformation is a new resource. Initially transformations will be user-selectable from a list of common operations, like conversion between standard and actual flow rate. Eventually, though, transformations should be user defined by typing the appropriate mathematical functions with a user defined function (UDF) language, much like mathematical transformations are user programmable in Excel.

By linking with electronic laboratory notebooks and other resource media, all documents, reports, and data generated for a particular experiment can be stored in one place. That is, disparate information, like weight measurements, spectroscopic analyses, and water quality analyses, can be connected to and accessible from the Adacta system. In this way, any measured or saved resource that is suspected in the future as relevant to the performance of a sample can be easily retrieved and evaluated. Then, if a device, like a mass flow controller, is considered suspect, its use in all test stands can be readily traced for further analysis by the scientists and engineers responsible for the test stand's operation.

Thus, by carefully archiving all associated data and detailed information about the test stand, a digital twin of the experimental setup (including all of the devices and samples) can be assembled. As the database of information grows, it will be possible to follow the history of catalysts, across reactor setups, to identify the smallest details that impact experimental performance, even far into the future when the experimental test stand is decommissioned, and key personnel have moved on to other roles. In fact, it is conceivable that



**Figure 11.** Simplified flowchart of the user interface that allows for quick transitions between Resource, Device, and Sample Views in Adacta.

eventually machine learning algorithms may be able to parse the vast amounts of data stored in an organization's Adacta repository to discover new insights.

The coupling of this RDM tool with numerical simulation tools is another exciting development underway. Conventionally, experimental data are summarized outside of a data historian to discrete values of velocity, composition, and so on, and these values are used by data input files for simulations that run outside of the experimental data system, with results that do not automatically synchronize with the data management system. Building on the CaRMeN [7, 8] technology, which allows for the rapid comparison of experimental data against simulation, Adacta will also be able to spawn and manage simulation results. In addition to interfacing with the DETCHEM software suite, Adacta will be able to eventually interface with any simulation software with the use of an appropriate driver.

Data security is another important area of development, because storing all important data in one place has the potential to create access issues to secrets and other information that may not be appropriate to disseminate even amongst a team. Eventually data sharing can be configured to expire after a period of time, to set editing permissions, and to hide information that the owner does not want to share with the recipient. Data sharing within an organization's RDM system will thus have many of the same options that data sharing on popular cloud-based file storage systems presently have, allowing for a high degree of control of what recipients get to see and edit within the database.

Summarizing, Adacta is a new RDM system that, although constructed with the experimental catalysis community in mind, is extendable to any type of experimental system where equipment records data in any format (data tables, photographs, videos, and so on). The key feature is traceability between resources (e.g., data and transformations), devices used to record data, and samples. Adacta records all the collected information internally as events on a timeline. This approach follows the natural course of how all devices and samples in a given system evolve over time, which ensures that all of the relevant information needed to create a time-accurate digital twin of the experimental apparatus is available to the research team. In the authors' experience, detailed information on an experimental setup is in most cases enough to be able to interpret and use the recorded data. Ultimately, as databases grow, organizations will be able to continually mine stored information for new insights, both to troubleshoot and to discover breakthroughs in relations that were previously masked and nearly impossible to find when data are stored in different systems and in different formats.

## Supporting Information

Supporting Information for this article can be found under DOI: <https://doi.org/10.1002/cite.202200064>.

The development of the concept for the software Adacta and its application at laboratory setups at KIT was supported by the NFDI4Cat project, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with the project number 670389-NFDI 2/1. Please contact the corresponding author for information on obtaining Adacta. The authors also acknowledge Dr. P. Lott, M. Borchers and N. Ruf for their contribution and valuable feedback during the development of the software. Additionally, S. Angeli and O. Deutschmann acknowledge funding by the TrackAct project, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 426888090 – SFB 1441 for the implementation of the RDM software. Open access funding enabled and organized by Projekt DEAL.

## References

- [1] T. P. Raptis, A. Passarella, M. Conti, *IEEE Access*. **2019**, *7*, 97052–97093. DOI: <https://doi.org/10.1109/ACCESS.2019.2929296>
- [2] M. D. Wilkinson et al., *Sci. Data* **2016**, *3* (1), 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>
- [3] R. Schlögl, *ChemCatChem*. **2017**, *9* (4), 533–541. DOI: <https://doi.org/10.1002/cctc.201700026>
- [4] C. Wulf et al., *ChemCatChem*. **2021**, *13* (14), 3223–3236. DOI: <https://doi.org/10.1002/cctc.202001974>
- [5] S. Weber et al., *ChemCatChem*. **2022**, *14* (8), e202101878. DOI: <https://doi.org/10.1002/cctc.202101878>
- [6] A. Trunschke et al., *Top. Catal.* **2020**, *63* (19–20), 1683–1699. DOI: <https://doi.org/10.1007/s11244-020-01380-2>
- [7] H. Gossler, L. Maier, S. Angeli, S. Tischer, O. Deutschmann, *Phys. Chem. Chem. Phys.* **2018**, *20* (16), 10857–10876. DOI: <https://doi.org/10.1039/c7cp07777g>
- [8] H. Gossler, L. Maier, S. Angeli, S. Tischer, O. Deutschmann, *Catalysts* **2019**, *9* (3), 227. DOI: <https://doi.org/10.3390/catal9030227>
- [9] O. Deutschmann et al., *DETCHEM Software package*, 2.8 ed., Karlsruhe **2020**. [www.detchem.com](http://www.detchem.com)
- [10] A. Nieva de la Hidalga, D. Decarolis, S. Xu, S. Matam, W. Yesid Hernández Enciso, J. Goodall, B. Matthews, C. R. A. Catlow, *Data Intell.* **2022**, *4* (2), 455–470. DOI: [https://doi.org/10.1162/dint\\_a\\_00143](https://doi.org/10.1162/dint_a_00143)
- [11] A. Nieva de la Hidalga, J. Goodall, C. Anyika, B. Matthews, C. R. A. Catlow, *Catal. Commun.* **2022**, *162*, 106384. DOI: <https://doi.org/10.1016/j.catcom.2021.106384>
- [12] *NFDI4Cat – NFDI for Catalysis-Related Sciences*, <https://nfdi4cat.org/> (Accessed on April 29, 2021)
- [13] *Generating knowledge with myhte - Data Management and Evaluation*, [www.hte-company.com/en/solutions/technology-solutions-ts/software-solutions/generating-knowledge](http://www.hte-company.com/en/solutions/technology-solutions-ts/software-solutions/generating-knowledge) (Accessed on May 13, 2021)
- [14] C. Fütter et al., in *Modern Applications of High Throughput R&D in Heterogeneous Catalysis* (Eds: A. Hagemeyer, A. F. Volpe), Bentham Science Publishers, Sharjah, U.A.E. **2014**.