# Nanoelectronics and Its New Materials – A NEW ERA OF NANOSCIENCE

**Nanoelectronics and Its New Materials –**

## A NEW ERA OF NANOSCIENCE

ZEISS
Seeing beyond

WILEY

**Discover the recent advances in electronics research and fundamental nanoscience.**

Nanotechnology has become the driving force behind breakthroughs in engineering, materials science, physics, chemistry, and biological sciences. In this compendium, we delve into a wide range of novel applications that highlight recent advances in electronics research and fundamental nanoscience. From surface analysis and defect detection to tailored optical functionality and transparent nanowire electrodes, this eBook covers key topics that will revolutionize the future of electronics.

To get your hands on this valuable resource and unleash the power of nanotechnology, simply download the eBook now. Stay ahead of the curve and embrace the future of electronics with nanoscience as your guide.

ZEISS

Seeing beyond

WILEY

# The Impact of Digitalized Data Management on Materials Systems Workflows

*Frank Tristram, Nicole Jung, Patrick Hodapp, Rasmus R. Schröder, Christof Wöll, and Stefan Bräse\**

The basic modules for materials research are systems for the design, synthesis, preparation, analysis, and application of materials and materials systems. To be efficient and produce findable, accessible, interoperable, and reusable (FAIR) data, state-of-the-art materials research needs to consider the integration of research data management (RDM) workflows and, in the end, the implementation of process automation concepts for all parts of the main modules. Here, the state-of-the-art methods of RDM in academia are described and a perspective on the future of digitalized molecular material systems workflows is given. The different elements of an integrated research data management strategy are described, and examples of automated processes are depicted. As such, the use of electronic lab notebooks for comprehensive documentation, the use of data-integration and data-conversion strategies, and the establishment of two platforms that enable the automated synthesis of chemical components for materials and the analysis of materials by electron microscopy, are highlighted. Two examples of beneficial effects of successful RDM strategies are presented, showing a sophisticated tool for data prediction based on machine learning and options for creating community-driven databases by extracting and re-using data from different scientific projects.

## 1. Introduction: Importance and Most Essential Components of Digitalization in Materials Research

Digitalization is rapidly becoming more and more important in most of the scientific disciplines and is also an essential part of the research processes in molecular materials research.[1] There are already manifold impressive examples showing how digitalization enables scientists in academia and industry to access and analyze vast amounts of data quickly and accurately. Sophisticated applications such as machine learning and other data analysis tools or specific software can drastically improve and accelerate scientific work. As a result, scientists make more informed decisions, and develop new materials faster and sometimes more economically.[2] These examples hint at the potential of a fully digitalized materials science, which has not yet come close to being achieved. Previous

F. Tristram
3DMM2O Cluster of Excellence (EXC-2082/1–390761711)
Karlsruhe Institute of Technology (KIT)
Schlossplatz 19, 76131 Karlsruhe, Germany

N. Jung, S. Bräse
Karlsruhe Nano Micro Facility (KNMFi)
Karlsruhe Institute of Technology (KIT)
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen,
Germany
E-mail: braese@kit.edu

N. Jung, P. Hodapp, S. Bräse
Institute of Biological and Chemical Systems–Functional Molecular
Systems (IBCS-FMS)
Karlsruhe Institute of Technology (KIT)
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen,
Germany

P. Hodapp
Institute for Biological Interfaces 3 – Soft Matter Laboratory
(IBG 3 – SML)
Karlsruhe Institute of Technology (KIT)
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen,
Germany

R. R. Schröder
BioQuant
Universität Heidelberg
Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

C. Wöll
Institute of Functional Interfaces (IFG)
Karlsruhe Institute of Technology (KIT)
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen,
Germany

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

examples each refer to a subarea of scientific work that is supported by digital processes, or they show work with a selection of digitalized data. Many hurdles still have to be overcome on the way to a comprehensive scientific digitalization, including the provision of technologies for digitalization and their systematic implementation. This is also important for the adaptation of previous working methods to the generation of findable, accessible, interoperable, and reusable (FAIR) data. In the end, the digitalization of scientific work needs to be supported by a sustainably hosted infrastructure that meets the requirements of the community and can be seamlessly integrated into the scientists' workflows.

In the following sections, we exemplarily describe components to achieve a high level of digitalization in materials sciences for the long term. Being aware of the diversity of existing or planned software, tools, infrastructure, and even standards that support digital scientific work in materials sciences, we selected components which can be highlighted with already existing implementations in a productive environment. Even though those components were established with a distinct focus on a discipline or method, the basic concepts behind the selected components and the models used for their implementation may be used to design a general approach toward an efficient digitalization strategy.

For materials science, as for many other disciplines, the recording, digital provision, and processing of data is a prerequisite to efficiently deal with digital data. Software such as electronic lab notebooks (ELNs) or laboratory information and management systems (LIMS) with the option to seamlessly integrate devices such as reactors, sensors, microscopes, or other analytical instruments are therefore an important pillar for digitalization in materials sciences. ELN and LIMS can further support the full research data life cycle including the collection, analysis, and preparation of data for a publication. ELNs that support discipline-specific standards and allow structured data storage can also be important software for the generation of machine-readable data (Chapter 2.1). With suitable technical instrumentation and robotics, this can pave the way to automation processes driven by digitalization, also minimizing the individual human factor for the repetition of experiments (Chapter 2.2.). The design of a research data management infrastructure meeting the needs of full digitalization can be based on manifold open source or free software and tools. Still, the success of digitalization strategies may depend on suitable models to include proprietary solutions for cases in which a high technological standard cannot be maintained with community-driven tools. This, for example, affects the wide area of electron microscopy that additionally is in need of a decentralized knowledge architecture (Chapter 2.3). The establishment of infrastructure, software, and tools that support a digitalized scientific work environment is a laborious effort that needs many decisions and probably adaptations of current workflows. Still, the benefits are enormous if data are available to be systematically analyzed and re-used. Data-driven approaches to materials research, such as machine learning and big data analysis, are key assets to accelerate scientific work. Examples are the identification of patterns that can be used to optimize materials performance or the prediction of synthetic parameters for optimized materials (Chapter 2.4.). The description of a modular infrastructure for digital research data management is completed with a concept for automated data annotation and conversion that helps to collect information in formats suitable for analysis

even if the data sources, file types, and standards are quite diverse (Chapter 2.5.).

For each use case, we provide a sketch of a workflow that may serve as an example template for the described research area.

## 2. The Parts of Digitalization in Materials Research

### 2.1. Electronic Lab Notebooks as the Foundation of Digital Workflows in Materials Research: Using ELNs to Access Data from Lab Devices

The digitalization of processes and documentation in the materials sciences depends, in particular, on the successful integration of data and scientific equipment into the digitalized research process. Only if a fully digitalized and loss-free handling of the data is realized, from the generation of the research data at the instrument to the storage and analysis of the data, can the potential benefits from digitalization be systematically exploited and established in the long term. Many areas of the industry have been systematically working on the digitalization of processes in the laboratory environment for many years, using systems such as LIMS, ELNs, and other (partially) automated solutions.[3–5] Especially, LIMS systems were used to connect instruments to digital laboratory environments and to digitalize the data flow completely without manual interruption. ELNs were used for the documentation of processes and results similar to the handwritten lab books. In the chemical industry, both were often used together and even integrated in the past. Nowadays the differentiation of ELNs and LIMS is quite difficult due to overlapping functions. Current ELN systems contain functionality of LIMS while systems called LIMS can have most of the functionality of an ELN.[6,7] In academic research environments and smaller companies, ELN and LIMS are mostly not available, due to cost. The use of devices with different measurement methods and from different manufacturers, as well as the generation of very different file formats, makes uniform control of processes and standardized data provision difficult. In recent years, however, it has been shown that systems for solving the current challenges of digitalization do not necessarily have to be proprietary and expensive. The development of open-source laboratory information software, such as open-source LIMS or ELNs, is a promising way to use efficient digitization methods even with limited financial resources.[6,8] Within the academic research groups of the cluster of excellence 3DMM2O[9] open-source solutions are used that allow a step-by-step integration of laboratory devices. The focus of device integration here is not only the transfer of data to centrally accessible storage resources and management systems, but a systematic integration of the generated data into the digital research process to realize the FAIR data principles wherever possible.[10] For this purpose, a combination of freely available methods, scripts, and open-source software is used. The solution within 3DMM2O is described here as an example for many other digitalization paths in the field of materials science, using free tools and software.

A first step toward laboratory digitalization can be the transfer of data, which is usually only locally available, from the measuring instruments to one or more central storage locations. This can be done systematically by using scripts that are installed locally on measuring computers and that manage the data transfer from measuring device to central storage locations (**Figure 1**, step 1).[11] An alternative is to set up additional

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

## Data flow from devices to electronic lab notebook
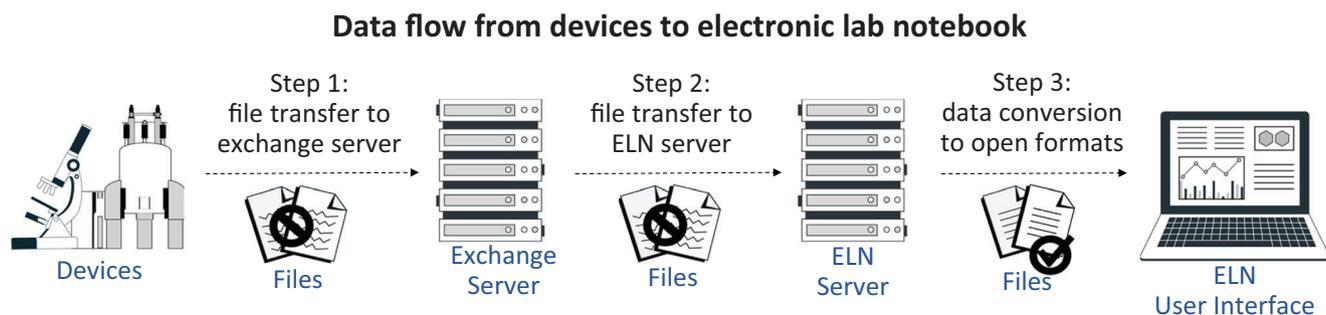


**Figure 1.** Schematic presentation of a simple workflow for data transfer from devices to an ELN and further processing to store machine and human-readable data. The process includes three main steps and has been realized as a data transfer concept in the discipline of chemistry with the Chemotion ELN. Images used for the figure were generated by C. Henken, Karlsruhe Institute of Technology (KIT), ZML - Center for Technology-Enhanced Learning, License: CC-BY.

memory locations in the software of the measuring devices, if they offer such an option.[12] Since the methods for data transfer always depend on the measuring device used, each device must be viewed individually to solve the process optimally. However, the users of the measuring devices can support each other and accelerate broad application by sharing the information through the already developed solutions. A list of such available methods has already been established for some of the devices used in materials science. Examples for such successful device integrations were shown for example for nuclear magnetic resonance (NMR) spectrometers, infrared (IR) spectrometers, or mass spectrometers.[13] This will be continuously expanded in the future with further examples and is open to the community for use or active contribution. Another important building block for the digitization of research processes and their documentation is the use of an ELN to enable the storage and linking of data, metadata, and other important information in the research process (Figure 1, step 2). It must be clearly stressed that digitalization of materials research requires as a first, essential step the use of an ELN to prepare experimental data and theoretical results for further processing.

The choice of a particular ELN depends largely on the functions required in each scientific discipline. In addition to the very well-known commercial solutions, various open source solutions have been established in recent years.[14] One advantage of the commercial solutions is the usually very high service level, while the advantages of open source tools are above all the possible adaptation by the community and thus their cost-effective extension. With wide distribution across research institutes, the basis for a gradual introduction of standardization elements such as uniform metadata schemas can be achieved. Another essential task of ELNs is the provision of measurement data via suitable user interfaces. For this purpose, queries can be implemented in ELNs that retrieve the centrally provided measurement data and assign them to the user. In the first of two typical examples of ELNs to be discussed here, the Chemotion ELN, these queries are performed at regular short time intervals, thus allowing easy integration of measurement data from NMR spectrometers, mass spectrometers, IR spectrometers, and many other instruments.[12] A special advantage of this implementation of Chemotion ELN is its flexibility with respect to the combination of different solutions for device integration, and thus independence from costly system integration. In the second example,

elabFTW,[15] a different strategy providing a higher flexibility as regards data types is pursued. For further data processing, ELNs can integrate additional tools or software that—if required—support the processing of open data, the extraction of metadata, and the generation of graphs for an easy and quick view of the data. The integration of subject-specific open-source software can promote independence from commercial software in many areas of work.[16] For work in molecular chemistry and materials characterization, developments such as NMRGlue, ChemSpectra,[17] NMRium,[18] LiberTEM,[19] and many others can enable work with analytical data without the use of external tools. The embedding of these libraries and tools allows the use of a variety of standardized file formats of commonly used measurement methods to be read and interpreted directly in the user interface.

In those cases where file formats are provided in a non-standardized but readable form, data converters[20] (see also Chapter 2.6) can provide the necessary standardization through an additional intermediate step (Figure 1, step 3). In this case, the task of the ELNs is a target-oriented combination of the methods and tools necessary and helpful for the corresponding communities, the efficient combination of the individual process steps, and the provision of all important data arising in the research data life cycle.

For the Chemotion application, this includes not only the storage of the various raw and processed files, but also a summary of the analytical results according to community standards and their checking for completeness and consistency. The digitalization of research processes thus allows research data to be digitally captured and made available in a readable, reusable form in an efficient workflow. In addition, the foundations are being laid for the further use of AI methods to provide scientists with the necessary tools for future-capable, high-performance scientific methods. Especially in the area of measurement data, the introduction of AI can support its evaluation and interpretation and discover errors, so that measurement data curation no longer has to be carried out by experts alone.

### 2.1.1. ELNs and Their Role in the Documentation of Scientific Work

In addition to providing measurement data and processing it, ELNs offer many other advantages that make digitalization in the materials sciences attractive to scientists and enable scientists to

## Benefits and characteristics of ELNs
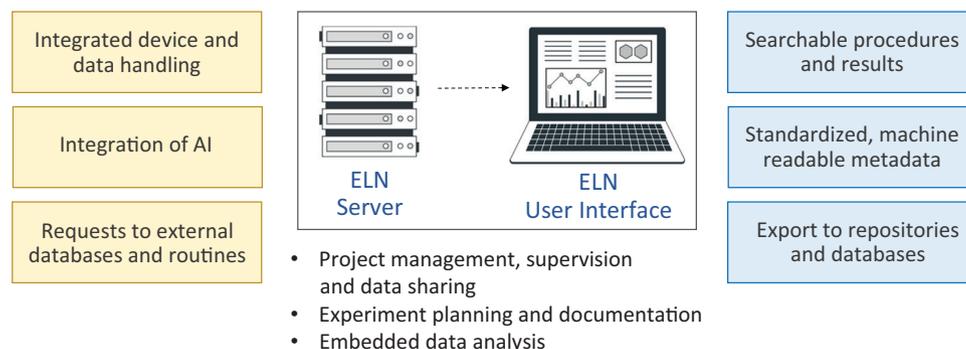


**Figure 2.** The benefits of using an ELN for the management, documentation, and analysis of scientific data. Left in yellow: benefits related to the support of the scientists' work processes; Right in blue: benefits related to a later re-use of the data and metadata. Images used for the figure were generated by C. Henken, Karlsruhe Institute of Technology (KIT), ZML - Center for Technology-Enhanced Learning, License: CC-BY.

work sustainably and accelerate their research (**Figure 2**). ELNs can very easily link the measurement data obtained with other information relevant to the research process. This includes the planning of experiments and description of the materials and processes they require, but also important administrative functions. Administrative functions allow data and information to be quickly added to the relevant documentation areas of the ELN, they allow ELN content to be shared with other partners, and they allow projects to be managed. Within the Cluster 3DMM2O, two ELNs (Chemotion-ELN[21] and elabFTW[15]) are in predominant use. While the generically oriented ELNs such as elabFTW allow free-text based documentation and are therefore highly flexible with respect to the content, style, and structure of the documentation, subject-specific ELNs or subject-supporting ELNs usually align the documentation with community standards. Those ELNs, such as Chemotion, are less flexible as they require certain data structures, but offer functions that support a broad range of research work in accord with the domain-specific needs of the scientists. For example, common data structures can be mapped and incorporated into the representation of scientific processes. Metadata can be queried using common metadata schemas,[22] and software for entering diverse content can standardize documentation by using a centrally available infrastructure. Examples in the Chemotion ELN software include the use of structure editors to enter molecular structures and their processing using common toolkits to generate database identifiers.[16,23]

Through such software and its integrated tools, standards can be introduced to, for example, search external databases such as PubChem.[24] Electronic laboratory notebooks can thus link laboratory documentation information systems provided by other vendors. In addition to database connectivity, ELNs can provide many other interfaces to systems of importance to researchers. For example, work with ELNs can be supported by connecting software for data extraction and reuse.

Examples from chemical synthesis include the integration of the software ChemScanner to extract data from external sources,[25] and the integration of systems to extract metadata from measurement data.[20] However, one of the most important functions of ELNs is the preparation of data for publication of results and in parallel with the disclosure of data in repositories.[26] The direct generation of publication-supporting materials such

as descriptions and texts, as well as the transfer of research data to repositories, is supported so far by a few ELNs relevant to materials science. In the long run, features like these can massively increase the digital availability of data and thus add real value for the scientific community. Examples of the need for repositories that provide suitable data for machine learning purposes were also seen in projects of the 3DMM2O cluster, where a lot of effort had to be invested in extracting data for metal–organic framework (MOF) reaction prediction (see Section 2.5). In order to be able to support other areas of materials science in addition to chemistry, the Chemotion ELN has been equipped with various functions that allow it to be adapted to other disciplines; this is known as the LabIMotion ELN (an extension of the Chemotion ELN).[27] Thus, a formerly domain-specific ELN can cover a broad range of scientific work and support interdisciplinary work. It is important for these developments to be supervised and supported intensively by the scientists in each case in order to be able to cover the requirements of the research tasks.

### 2.2. An Automated Platform for Efficient Syntheses in Molecular Materials Research

The automation of repetitive tasks in production and manufacturing has increased production output and improved the overall quality and consistency of results. Similarly, in industrial and academic laboratories, robotic automation of repetitive tasks eliminates human error and increases productivity, while allowing humans to dedicate their time to more challenging and not (yet) automatable tasks. Most of the laboratory automation efforts in research environments, especially in synthetic chemistry, have largely focused on automating specific steps—usually the ones that are repetitive, tedious, time-consuming, or most prone to human error, such as automation of sample preparation in analytical laboratories or flash column chromatography in organic chemistry. Some systems also allow the automation of one specific reaction or reaction type from preparation to purification, such as in flow chemistry setups. However, few projects have aimed to create a robotic synthetic research laboratory allowing for a wide spectrum of reactions with very different parameters and requirements.[28–30] Those projects commonly work at

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

microliter scales due to economic and environmental burdens that are connected to the use of large volumes and amounts of solvents and chemicals. Also, storage and handling are facilitated by adhering to small-scale industry standards, such as well-plate formats. For some research fields, nanoliter-scale reactions yield enough material, since modern analytical tools no longer require large amounts of product for biochemical analysis. However, for materials applications and prototype production, small amounts of chemical substances are not sufficient and much larger reaction volumes become necessary. Only a very few projects have attempted to replicate the complete batch chemistry workflow of organic chemical synthesis laboratories on larger reaction scales.[31,32]

The chemical analysis and synthesis automation platform (ChemASAP), under development at the time this article was written,[33] is a multipurpose synthesis, analysis, and characterization platform that attempts to fully map the workflow of organic chemical research laboratories in multi-milliliter–scale batch reactions. Based on a modular design, the platform allows for easy incorporation of new modules into hardware and software, thereby expanding its capabilities. Its flexible architecture allows the platform to evolve and adapt to fulfill new research targets and permit future integration of new technologies without the need to re-engineer the entire platform. ChemASAP covers both the software and hardware aspects of the platform.

The software is composed of modules, each one in charge of one specific task, such as the interpreter, which translates the chemist's instructions to ChemASAP internal commands and to specific tasks for each hardware component; the scheduler, for planning ahead of time which hardware components to send what task at what time; or the plugins, that interface the ChemASAP software with the hardware component executing the task. The hardware components, or modules, are partly developed in-house and partly commercially acquired. Simple hardware might only require the plugin to send the instructions provided by the scheduler via transmission control protocol/internet protocol to the device controller; however, more complicated modules, such as liquid chromatography/mass spectrometry (LC/MS), might require the plugin to perform more complicated tasks, such as controlling the vendor's proprietary software, monitoring the device's status, providing the device's software with the necessary information and commands, and exporting the data. On the hardware side, ChemASAP is also built in a modular way, with one module for every step or task in the workflow. Examples of the hardware modules to be included for a comprehensive approach are the robotic storage system for all chemicals and consumables; dosing platforms for dosing of solvents and liquid and solid chemicals; solvent evaporation modules;[34] a reaction station allowing for stirring, cooling, and heating of reaction container workup modules for filtration and liquid–liquid extraction; and purification modules, namely HPLCs with multiple detectors and mass spectrometers. A platform for chemical synthesis and analysis also needs chemical analytic and characterization such as gas chromatography-mass spectrometry, liquid chromatography-mass spectrometry, NMR, and UV–vis spectroscopy. Finally, a conveyor system needs to interconnect all modules to transport the chemicals, reaction vessels, and consumables while allowing easy integration of new modules. Many of the required modules can be ac-

quired from commercial vendors such as Zinsser Analytics,[35] Shimadzu,[36,37] Axel Semrau,[38] Biotage,[34,39] Beckhoff,[40] and many others. These offer very precise and specific devices for the required processes, but there are also some processes for which no suitable solution is available, to our knowledge. The latter devices have to be developed in house. An important factor for the success of automated platforms such as ChemASAP is a concept by which to integrate the platforms into digital research data management workflows. Ideally, the latter are the ones that are already accepted by the scientists. For the planning of chemical reactions with ChemASAP, software such as the Chemotion ELN can be used. The use of an ELN offers a lot of benefits, such as the design of experiments by individual scientists within the work environment they are familiar with. A main achievement that enables the use of ELNs as planning software for automated systems is the generation of machine-readable data. This and many other challenges are currently elaborated at KIT Karlsruhe in different projects. The link of Chemotion ELN to ChemASAP is planned to be finished in 2024. The combination of both will offer chemists an all-in-one solution for reaction planning, execution, analysis and characterization, data acquisition, storage, visualization, evaluation, and even the reporting of processes. Reactions will be submitted to the ChemASAP platform via the "reaction-ui" currently under development as part of Chemotion, allowing the chemist to utilize said to create and submit the reaction to be performed by the ChemASAP platform. All data, metadata, and parameters recorded during the experiment run, including data recorded by the analytical devices of ChemASAP will be sent to the user's Chemotion ELN for interpretation and scientific evaluation. The long-term aim here is to automatically abide by modern standards, such as the FAIR data principles, with no additional effort. The process data that are produced with ChemASAP, including all experimental parameters, metadata, and information, which are not recorded in normal experimental setups in chemical research labs, will be available in machine-readable form. Data from projects such as ChemASAP might therefore be highly valuable when trying to extract information from databases for machine learning (ML) purposes or for the development of AI algorithms. In this context, automated platforms such as ChemASAP could be further developed to allow for rapid unsupervised experimental testing and feedback loops to the algorithms driving it in the future. The combination of AI, simulation, and experimental automation is enabling a new era of materials discovery, as materials can be discovered and optimized faster and more efficiently than ever before.[41] Such an approach to materials discovery could open up new possibilities for the development of novel materials with improved properties, while enabling completely new applications and technologies.[42–44] The components of ChemASAP will be developed as open source software and open hardware components wherever possible to enable further re-use of its components in other institutions.

## 2.3. Automated Data Acquisition and Analysis Might Change Materials Research—Electron Microscopical Data as an Example

While intuitive research and disruptive approaches will always be essential parts of scientific progress, it is helpful to consider the

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

approach to big data and automated analysis workflows in another, recently flourishing field—cryo-electron microscopy and macromolecular imaging. Once basic physical rules for sample preparation, data collection, and data processing were defined in the mid-1990s and the following decade,[45] it was obvious that intensive data processing of massive numbers of images would deliver rich results. Therefore, database-driven microscope automation (e.g., Leginon[46,47]) and processing pipelines (e.g., Appion[48]) were developed and used extensively. Even though further advances in acquisition and processing software did at first not happen in a centralized or coordinated way (e.g., xmipp,[49] EMAN,[50] RELION,[51,52] and CryoSPARC[53,54]), the community realized the advantage of handling images in databases and going from there to automated image-processing workflows and 3D reconstructions. And, while early 3D reconstruction algorithms and processing pipelines relied heavily on the intervention of a person, situationally handling interdependent processing cycles, this setting has changed substantially. About 10 years ago the application of Bayesian approaches—and massive computing power to crunch large data sets—allowed 3D reconstruction to also be completely automated.[51] Today, for automated electron microscopes, companies provide processing of the image data stream to yield a 3D structure that is updated in real time to accommodate newly acquired data.[55]

The final results are then archived in centralized community repositories, such as a database for the raw images (EMPIAR[56]) and the final reconstructions and models (EMDB[57]).

How does this compare with, for example, the 3D imaging and materials characterization of 3D printed carbon materials at the nanoscale? Here, the situation is not yet anywhere close to the scenario discussed above. And, in our opinion, there will hardly be one unified processing workflow for all microscopical modalities such as imaging or spectroscopy of large volumes and at the nanoscale. However, the sheer number of images necessary for a statistically validated material characterization[58]—for example, by electron energy loss spectroscopy (EELS)—can be efficiently handled only in a database structure administered by an ELN/LIMS. Thus, microscopical data acquisition, data storage, data processing, final data, and result archiving are connected in a processing pipeline that has a modular structure and is operated from the ELN/LIMS. In the wider community, separate scripting of these individual steps is common, but microscope developers are pushing into an operating regime which reduces instruments to black boxes that run fully automated and also provide metadata directly to the processing software. An example is the metadata transfer between the TEM automation software Serial EM[59] and RELION.[51]

Especially, analytical characterization techniques such as EELS will become more prevalent as soon as large data sets allow one to show not only the "typical example" but also statistically sound results. This will come with large numbers of samples, large amounts of spectroscopic data, and automated processing pipelines. The situation will expand dramatically as soon as machine learning and Bayesian inference for the actual design of experiments at the microscope are in more widespread use. In the microscope, sample areas will be detected automatically, from one area of the sample for data recording to the next in an autonomous way.[60] At the same time, online processing will predict better acquisition parameters for the next sample area,

adapting acquisition to, for example, sample thickness or materials automatically detected in the new sample area. With these developments, which are being pursued in many labs, electron microscopy in the field of materials science will adopt many of the automation steps described for cryo EM, but add the most modern statistical and machine learning approaches to the more complex microscopy of materials samples. In the end, materials electron microscopy has the chance to develop into a high-throughput technique complemented by statistical predictions of effects at the nanoscale that are relevant to material properties and fabrication.

What are the data processing needs resulting from these high-throughput approaches in materials science? What is currently being developed is a processing environment that allows a very flexible, modular pipeline design for 1) experimenting with different algorithms or parameters; and 2) repetitive processing of large numbers of data sets (in production mode). This might be unified in a larger repository structure such as is currently developed in the NOMAD Laboratory structure,[61] or it can be organized locally. One must understand that the centralized structures currently in use in structural biology originated at a time, when individual institutions could i) not afford large scale data storage, and ii) possible concerns connected to individual data ownership and the possibility of decentralized FAIR data availability was not anticipated. The advantages of the centralized repository solution were too evident: A centralized solution will always provide a single, well-defined structure of metadata including a community-wide ontology.

Today, the awareness of data ownership has changed. Some studies want to provide access to more data than they are asked for by centralized repositories, they might want to restrict access to such additional data individually, and institutions might want to analyze data access themselves. Decentralized FAIR data availability, including therefore, for example, data processing scripts, seem equally favorable. However, this poses the important question of the compatibility of data documentation, scripting, and publishing between different decentralized repositories. Thus, the definition of metadata comes into focus, and at present several efforts across different electron microscopy laboratories are engaged in providing such a metadata scaffold for the community (e.g., the Helmholtz metadata collaboration—HMC[62] with the Matter–Hub focusing on metadata from electron microscopy[63]).

The situation described above might get even more complex. Structured research programs often combine studies in different fields of research—possibly each with their own centralized data repositories, or using local and public data storages such as Zenodo,[64] Github,[65] or Omero.[66] In general, then there is no common ontology, and meta-analysis of data years after the original deposition of these data becomes increasingly difficult.

However, such collaborative research activities are more and more a scientific necessity and have, in the meantime, certainly proven their benefits. In the traditional setting, this would imply that each individual scientific community has its own repositories, which are completely independent from each other. The problems with this scenario might be overcome by an open access to a metadata collection from all—centralized and decentralized—repositories (see **Figure 3**). Queries from and to this metadata server will create a unified,

## Metadata server: Including specific metadata into a data infrastructure



mutual harvesting

Community Exchange Servers

Files

Metadata Server

mapping

Files

ELN Server

research case specific information

Files

ELN User Interface

FAIR data publishing
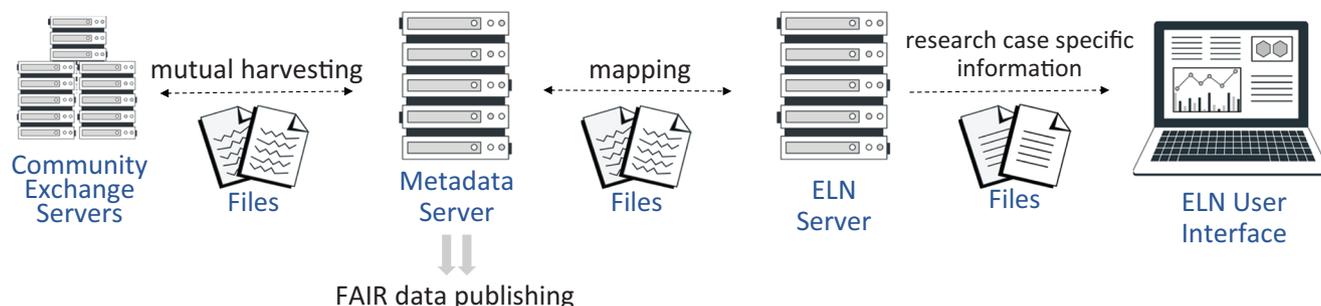
**Figure 3.** Schematic presentation of a workflow for data transfer from decentralized sources, contributing to one research topic. A joined metadata server handles all data and metadata requests and serves as a gatekeeper from and for the outside world. Images used for the figure were generated by C. Henken, Karlsruhe Institute of Technology (KIT), ZML - Center for Technology-Enhanced Learning, License: CC-BY.

repository-like experience, making the underlying individual and heterogeneous—centralized and decentralized—storage structure easily accessible.

### 2.4. Data Prediction and Analysis with Artificial Intelligence: MOFs and SURMOFs as an Example of the Application of Data Science to Materials Discovery and Optimization

MOFs are a very rich class of materials: the number of structurally characterized compounds listed in the Cambridge Structural Database just exceeded 100 000.[67] Since the number of possible applications of MOFs is rapidly increasing, there is a strong urge to create new variants of these reticular frameworks. A number of theoretical approaches have been established to predict new types of MOFs (in some papers, more than 2 million MOFs have been predicted). Recent work by B. Smit and coworkers has shown how artificial intelligence and machine learning can be applied to conquer this huge chemical space and to predict compounds with optimized properties.[68,69] Despite the progress in optimizing properties of MOFs in silico, the synthesis of such compounds is much more difficult to predict. Precise reaction conditions, concentrations, temperatures, and reaction times have to be determined in rather laborious trial-and-error approaches. In this context, the work of Tsotsalas and coworkers seeks for improvement.[70] For a given new, not-yet-synthesized MOF with a structure predicted using computational tools, they first carried out a literature search describing the synthesis of similar compounds. Then, for the publications identified in this search, they used natural language processing (NLP) to extract the synthesis conditions. These NLP results were then used to derive predictions of synthesis parameters for the new compound. Several test cases revealed that although the predictions were not perfect, the performances of ML-algorithms on that basis were already much better than that of human experts who provided their expectations. In this study, the time-limiting step was the NLP analysis, which was hampered by different authors using different terms to describe reaction conditions. If the MOF synthesis data were contained in a repository like Chemotion, the collection of suitable data for machine learning (ML) models would have been much faster. This result is a clear demonstration of the

need for an efficient, standardized storage of synthesis parameters for MOFs (see **Figure 4**).

The huge potential of AI methods can be demonstrated using a second example, concerning MOF thin films. For a number of advanced applications, for example in the context of electronic and sensor devices, MOFs need to be deposited in the form of thin films, with a high degree of orientation and low defect density. A very successful method in this context, layer-by-layer deposition, works in a kinetically controlled regime and thus requires a careful, time-consuming optimization of condition parameters. Since the number of these parameters (concentrations, temperature, and substrate immersion times) is large, their optimization is rather time-consuming. In a recent paper, Tsotsalas and coworkers demonstrated that ML methods are suited to greatly accelerate this optimization.[71] Interestingly, using genetic algorithms, a fully automated and optimized growth of a highly ordered, low-defect monolithic MOF thin film with an unusual orientation could be obtained. An analysis of the resulting synthesis conditions was surprising: for example, the ML algorithm identified a huge difference in optimum synthesis concentrations (two orders of magnitude), in contrast to the common belief that reactant concentrations should be similar.

### 2.5. Data Reuse

The promise of linked data and the semantic web has been around for decades. But in most cases rich and FAIR information is not yet as readily available as the data. Of course, there are plenty of easy-to-use repositories today that provide valuable data and metadata. A collection of such repositories and databases can be retrieved from re3data.org—Registry of Research Data Repositories.[72] but these are often only used within a discipline for common research objects like stars, weather, atoms, or molecules that are distinctive by nature. Comparing and combining information from designed scientific setups or composite structures with unique components is a much more tedious endeavor.

It is obviously easier to agree on metadata and file formats connected to a widely accepted object than on something that was built and is being used in only one lab. This is a major

## (Automated) data generation and handling for machine learning
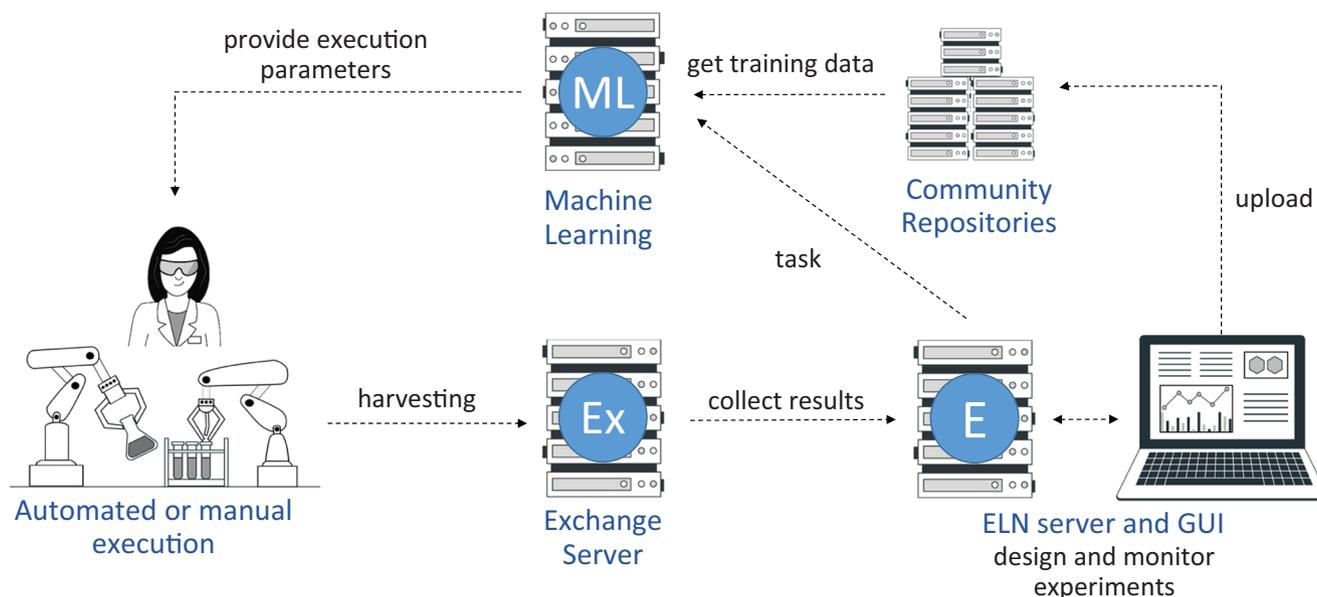


**Figure 4.** Proposed workflow for the integration of external knowledge harvesters and internal data sources to enable AI methods. Images used for the figure were generated by C. Henken, Karlsruhe Institute of Technology (KIT), ZML - Center for Technology-Enhanced Learning, License: CC-BY.

obstacle to information retrieval. Within a typical materials science

environment with a plethora of different instruments and research questions, there is a need for better data flow than downloading .dat files or copying tables or numbers manually or via single-use self-written scripts.

There are two possible paths that could lead directly to the availability of FAIR file formats in the future:

1) Data are converted by dedicated converters into standardized open file formats that can be re-used with high coverage of the original content due to the adaptation of converters to specific challenges. This FAIRification is most efficient if it is done directly after the data are generated.
2) Data are read, understood, and converted by generic data readers developed to deal with the majority of file formats and that are perhaps less efficient in capturing details for special data and metadata.

We have investigated both paths. A solution toward data conversion of type 1 is described as part of Section 2.2 of this perspective. Converters of type 1 can support research data management workflows to convert known data files into open standardized files with comprehensive coverage of the included information.

The second path is even more challenging, but necessary if we ever want to provide a FAIR future for "cold" data; for example, all files on the generic data repository zenodo.[73] Most scientific formats have not been invented for automated information harvesting to fill databases and repositories. Most formats have been and will be around for more than 10 years and are not going to change. As a first step in dealing with hundreds of different for-

mats, we currently use self-written software that can extract most information from any ASCII file (e.g., measurements or input or output from simulations) and can convert it into a technically standardized HDF5. Such a program does something best described as file segmentation. The task is to identify blocks and relations in a file along with encoding, data types, and typical patterns. Such disassembled file content can then, in a second standardization step, be analyzed and ingested into databases that can fulfill the semantic promises mentioned above.

To investigate their relevance and functionality for materials science, we downloaded 10 000 .dat, .txt, .json, and .csv files that zenodo ranked highest when given a materials science search string we derived from Wikipedia subcategory names. From 1605 included .dat files for which we could identify an encoding, we counted 122 different underlying formats with the all2HDF5 converter prototype mentioned above.

This only states the problem's size without solving any practical aspect of it, yet, because the semantic of all these files is different, too. However, now all the information is machine-readable at our fingertips and we are seeking for a future interoperability that leaves behind the barriers between formats, data creators, and subdomain standards. Especially for larger collaborations, such interoperability will reduce the development cycle time and the traceability of effects and results between groups. Embedding information, not just data, for all target groups and their different questions, equipment, and knowledge levels can probably never succeed perfectly. But automated segmentation and mapping between file description and segments, keys and values, and structural elements and tables mimicking human reading promise at least unprecedented flexibility and speed, transforming the coding work for data ingestion into a simple configuration. This might even allow us to harvest data graveyards in a

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de
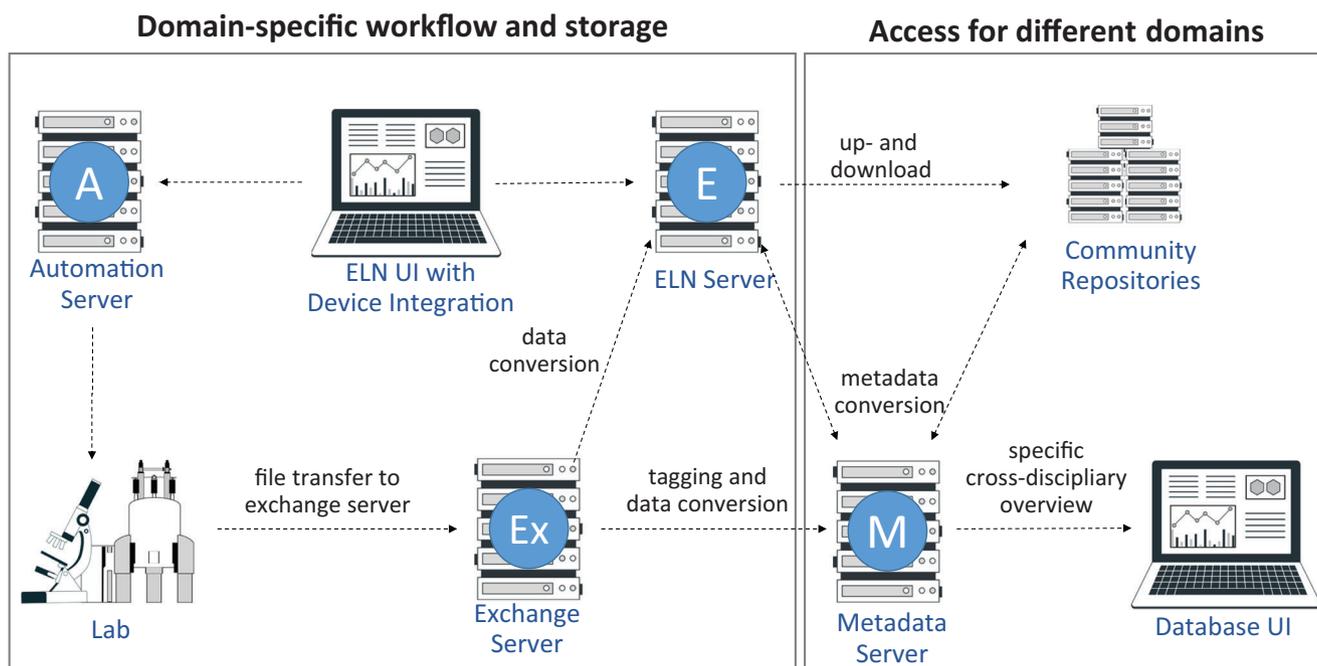
**Figure 5.** Developing the metadata server toward a local information platform that tries to gather and present all available information. Images used for the figure were generated by C. Henken, Karlsruhe Institute of Technology (KIT), ZML - Center for Technology-Enhanced Learning, License: CC-BY.

meaningful manner. Only such an easy-to-reach FAIRification can push comparative work, reuse of data, and the combination of information beyond the limits of individual laboratories. In fact, we need standard tasks like foreign data importation and comparison to be achieved without coding or tedious manual conversion steps, to raise quality standards and increase our opportunities to survey a broad field (**Figure 5**).

## 3. Conclusion

To achieve a completely digital research data management, it is necessary to integrate all aspects of the research process, from the initial idea to the final results. This includes the whole research data lifecycle consisting of the collection and storage of data, the analysis of data, the sharing of data, and the archiving of FAIR data. Currently, different initiatives such as the national research data infrastructure (NFDI)[74] and distinct Helmholtz programs develop and establish valuable concepts and software to support the cultural change toward a digitalized research environment in Germany. Achievements of these community developments were combined with data-driven workflows and projects within the cluster of excellence 3DMM2O to show how successful collaborative work is being tackled on the local level. To digitally exploit our collective knowledge across domains is the next big challenge for materials science. There is no end in sight to where this progress in computer technology will lead. What materials scientists did laboriously by hand 20 years ago can now be done at the push of a button. It is to be expected that what today can be done laboriously by hand, for example, in inverse material design, will also be possible in 20 years practically at the push of a button. However, the sheer endless abundance of materials and their surprises will for sure not have been worked

off to any extent in 20 years. We can already see that computer programs are capable of discovering unexpected properties, materials, and experimental procedures. Doing this at the push of a button is hardly imaginable for us yet, but maybe will come thereafter.

## Conflict of Interest

The authors declare no conflict of interest.

[1] S. Lazzari, A. Lischewski, Y. Orlov, P. Deglmann, A. Daiss, E. Schreiner, H. Vale, in *Advances in Chemical Engineering* (Eds: D.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

Moscatelli, M. Sponchioni), Academic Press, Cambridge, MA **2020**, pp. 187–227.

[2] V. Anipa, A. Tarzia, K. E. Jelfs, E. V. Alexandrov, M. A. Addicoat, *R. Soc. Open Sci.* **2023**, *10*, 220813.

[3] G. A. Gibbon, *Lab. Autom. Inf. Manage.* **1996**, *32*, 1.

[4] D. O. Skobelev, T. M. Zaytseva, A. D. Kozlov, V. L. Perepelitsa, A. S. Makarova, *Meas. Tech.* **2011**, *53*, 1182.

[5] PerkinElmer_TheNextFrontier, https://perkinelmerinformatics.com/sites/default/files/2022-12/PerkinElmer_TheNextFrontier_ebook_final.pdf (accessed: August 2023).

[6] C. Barillari, D. S. M. Ottoz, J. M. Fuentes-Serna, C. Ramakrishnan, B. Rinn, F. Rudolf, *Bioinformatics* **2016**, *32*, 638.

[7] K. Davies, *GEN Edge* **2020**, *2*, 303.

[8] Wiley Analytical Science: LIMS im Zeitalter von Open Source, https://analyticalscience.wiley.com/do/10.1002/was.00050095/ (accessed: May 2023).

[9] Cluster Office, 3D matter made to order (3DMM2O) – homepage, Cluster Office, 2023, https://www.3dmattermadetoorder.kit.edu/ (accessed: August 2023).

[10] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, et al., *Sci. Data* **2016**, *3*, 160018.

[11] M. Starman, ELN_file_watcher, ComPlat – Compound Platform @ Karlsruhe Institute of Technology (KIT), 2022, https://github.com/ComPlat/ELN_file_watcher (accessed: August 2023).

[12] J. Potthoff, P. Tremouilhac, P. Hodapp, B. Neumair, S. Bräse, N. Jung, *Anal. Chim. Acta: X* **2019**, *1*, 100007.

[13] List of enabled devices, https://www.chemotion.net/docs/eln/devices/list (accessed: August 2023).

[14] M. Branschofsky, D. Chudnov, in Proceedings Of the Second ACM/IEEE-CS Joint Conf. On Digital Libraries – Jcdl '02, ACM Press, New York **2002**.

[15] N. CARPi, A. Minges, M. Piel, *J. Open Source Softw.* **2017**, *2*, 146.

[16] S. Kotov, P. Tremouilhac, N. Jung, S. Bräse, *J. Cheminform.* **2018**, *10*, 38.

[17] Y.-C. Huang, P. Tremouilhac, A. Nguyen, N. Jung, S. Bräse, *J. Cheminform.* **2021**, *13*, 8.

[18] NMRium – Visualize and process NMR spectra, https://www.nmrium.org/ (accessed: February 2023).

[19] A. Clausen, D. Weber, K. Ruzaeva, V. Migunov, A. Baburajan, A. Bahuleyan, J. Caron, R. Chandra, S. Halder, M. Nord, K. Müller-Caspary, R. Dunin-Borkowski, *J. Open Source Softw.* **2020**, *5*, 2006.

[20] The chemotion data and metadata converter, https://github.com/ComPlat/chemotion-converter-app (accessed: August 2023).

[21] P. Tremouilhac, A. Nguyen, Y.-C. Huang, S. Kotov, D. S. Lütjohann, F. Hübsch, N. Jung, S. Bräse, *J. Cheminform.* **2017**, *9*, 54.

[22] Chemotion-Templates: Repository for collecting Polymer Templates for the Chemotion group, https://github.com/ComPlat/Chemotion-Templates, Github (accessed: August 2023).

[23] RDKit, https://www.rdkit.org (accessed: May 2023).

[24] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton, *Nucleic Acids Res.* **2023**, *51*, D1373.

[25] A. Nguyen, Y.-C. Huang, P. Tremouilhac, N. Jung, S. Bräse, *J. Cheminform.* **2019**, *11*, 77.

[26] P. Tremouilhac, P. C. Huang, C. L. Lin, *Chem.: Methods* **2021**, *1*, 8.

[27] The LabIMotion extension, https://chemotion.net/chemotionsaurus/docs/eln/labimotion (accessed: January 2023).

[28] Eli Lilly – robotic cloud lab, https://investor.lilly.com/news-releases/news-release-details/eli-lilly-and-company-collaboration-strateos-inc-launch-remote (accessed: February 2023).

[29] A. B. Santanilla, E. L. Regalado, T. Pereira, M. Shevlin, K. Bateman, L.-C. Campeau, J. Schneeweis, S. Berritt, Z.-C. Shi, P. Nantermet, Y. Liu, R. Helmy, C. J. Welch, P. Vachal, I. W. Davies, T. Cernak, S. D. Dreher, *Science* **2015**, *347*, 49.

[30] D. Perera, J. W. Tucker, S. Brahmbhatt, C. J. Helal, A. Chong, W. Farrell, P. Richardson, N. W. Sach, *Science* **2018**, *359*, 429.

[31] S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, L. Cronin, *Science* **2019**, *363*, https://doi.org/10.1126/science.aav2211.

[32] IBM RoboRXN, https://research.ibm.com/science/ibm-roborxn/ (accessed: February 2023).

[33] I. (s.). Oberdorf, KIT – KIT – media – press releases – archive press releases – automated chemical synthesis: Reliable production and rapid knowledge gain, Oberdorf, Iris (SEK), https://www.kit.edu/kit/english/pi_2021_072_automated-chemical-synthesis-reliable-production-and-rapid-knowledge-gain.php (accessed: August 2023).

[34] Biotage, Biotage – Evaporation & concentration, https://www.biotage.com/evaporation (accessed: February 2023).

[35] Zinsser Analytics, https://www.gardnerdenver.com/de-de/zinsseranalytic/ (accessed: February 2023).

[36] R. K. Hofstetter, M. Hasan, C. Eckert, A. Link, *ChemTexts* **2019**, *5*, 13.

[37] K. Shinada, K. Kitano, US Patent **2018**, US 10,161,906 B2.

[38] Axel Semrau, https://www.axelsemrau.de/das-unternehmen/hersteller/ctc-analytics (accessed: March 2023).

[39] W. Neil, G. Zipp, G. Nemeth, M. F. Russo, D. S. Nirschl, *SLAS Technol.* **2018**, *23*, 412.

[40] Beckhoff Automation GmbH & Co. KG, Hülshorstweg 20, 33415 Verl, Germany, [No title], https://www.beckhoff.com/de-de/produkte/motion/xplanar-planarmotorantriebssystem/?pk_campaign=AdWords-AdWordsSearchxplanar_de&pk_kwd=xplanar (accessed: August 2023).

[41] M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. M Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu, A. Aspuru-Guzik, *Acc. Chem. Res.* **2022**, *55*, 2454.

[42] G. B. Olson, *Science* **2000**, *288*, 993.

[43] J. Cai, X. Chu, K. Xu, H. Li, J. Wei, *Nanoscale Adv.* **2020**, *2*, 3115.

[44] N. Yoshikawa, A. Z. Li, K. Darvish, Y. Zhao, H. Xu, A. Aspuru-Guzik, A. Garg, F. Shkurti, arXiv:2212.09672v1 **2022**.

[45] advanced-chemistryprize2017-1.pdf. The Development of Cryo-Electron Microscopy, Scientific Background on the Novel Prize in Chemistry 2017 (accessed: August 2023).

[46] A. Cheng, C. Negro, J. F. Bruhn, W. J. Rice, S. Dallakyan, E. T. Eng, D. G. Waterman, C. S. Potter, B. Carragher, *Protein Sci.* **2021**, *30*, 136.

[47] D. Fellmann, J. Pulokas, R. A. Milligan, B. Carragher, C. S. Potter, *J. Struct. Biol.* **2002**, *137*, 273.

[48] G. C. Lander, S. M. Stagg, N. R. Voss, A. Cheng, D. Fellmann, J. Pulokas, C. Yoshioka, C. Irving, A. Mulder, P.-W. Lau, D. Lyumkis, C. S. Potter, B. Carragher, *J. Struct. Biol.* **2009**, *166*, 95.

[49] J. M. de la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldívar, J. Vargas, J. M. Carazo, C. O. S. Sorzano, *J. Struct. Biol.* **2013**, *184*, 321.

[50] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, S. J. Ludtke, *J. Struct. Biol.* **2007**, *157*, 38.

[51] D. Kimanius, L. Dong, G. Sharov, T. Nakane, S. H. W. Scheres, *Biochem. J* **2021**, *478*, 4169.

[52] S. H. W. Scheres, *J. Struct. Biol.* **2012**, *180*, 519.

[53] A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, *Nat. Methods* **2017**, *14*, 290.

[54] Structura Biotechnology, https://structura.bio (accessed: March 2023).

[55] Real-time Cryo-EM Software Solutions, https://cryosparc.com/live (accessed: March 2023).

[56] A. Iudin, P. K. Korir, S. Somasundharam, S. Weyand, C. Cattavitello, N. Fonseca, O. Salih, G. J. Kleywegt, A. Patwardhan, *Nucleic Acids Res.* **2023**, *51*, D1503.

[57] EMDB, Electron microscopy data bank (EMDB), https://www.ebi.ac.uk/emdb/ (accessed: March 2023).

[58] J. A. Kammerer, F. Feist, D. Ryklin, A. Sarkar, C. Barner-Kowollik, R. R. Schröder, *Adv. Mater.* **2023**, *35*, 2211074.

[59] D. N. Mastronarde, *J. Struct. Biol.* **2005**, *152*, 36.

[60] N. Creange, O. Dyck, R. K. Vasudevan, M. Ziatdinov, S. V. Kalinin, *Mach. Learn.: Sci. Technol.* **2022**, *3*, 015024.

[61] Mission – NOMAD lab, https://nomad-lab.eu (accessed: April 2023).

[62] The Helmholtz Metadata Collaboration (HMC), https://helmholtz-metadaten.de/en (accessed: April 2023).

[63] Matter-Hub focusing on Metadata from electron microscopy, https://helmholtz-metadaten.de/en/matter/uebersicht (accessed: April 2023).

[64] The repository Zenodo, https://zenodo.org/ (accessed: May 2023).

[65] The code repository Github, www.github.com (accessed: May 2023).

[66] Omero, the open microscopy environment, https://www.openmicroscopy.org/omero/ (accessed: April 2023).

[67] R. Freund, S. Canossa, S. M. Cohen, W. Yan, H. Deng, V. Guillerm, M. Eddaoudi, D. G. Madden, D. Fairen-Jimenez, H. Lyu, L. K. Macreadie, Z. Ji, Y. Zhang, B. Wang, F. Haase, C. Wöll, O. Zaremba, J. Andreo, S. Wuttke, C. S. Diercks, *Angew. Chem. Int. Ed Engl.* **2021**, *60*, 23946.

[68] S. M. Moosavi, B. Á. Novotny, D. Ongari, E. Moubarak, M. Asgari, Ö. Kadioglu, C. Charalambous, A. Ortega-Guerrero, A. H. Farmahini, L. Sarkisov, S. Garcia, F. Noé, B. Smit, *Nat. Mater.* **2022**, *21*, 1419.

[69] Y. Kang, H. Park, B. Smit, J. Kim, *Nat. Mach. Intell.* **2023**, *5*, 309.

[70] Y. Luo, S. Bag, O. Zaremba, A. Cierpka, J. Andreo, S. Wuttke, P. Friederich, M. Tsotsalas, *Angew. Chem. Int. Ed Engl.* **2022**, *61*, e202200242.

[71] L. Pilz, C. Natzeck, J. Wohlgemuth, N. Scheuermann, P. G. Weidler, I. Wagner, C. Wöll, M. Tsotsalas, *Adv. Mater. Interfaces* **2022**, *10*, 2201771.

[72] The repository re3data, https://www.re3data.org/ (accessed: May 2023).

[73] Zenodo is a general-purpose open repository that is operated by CERN and can be used for the deposition of a variety of digital research outcome.

[74] N. Hartl, E. Wössner, Y. Sure-Vetter, *Inf. Spektrum* **2021**, *44*, 370.

[75] C. Steinbeck, O. Koepler, F. Bach, S. Herres-Pawlis, N. Jung, J. C. Liermann, S. Neumann, M. Razum, C. Baldauf, F. Biedermann, T. W. Bocklitz, F. Boehm, F. Broda, P. Czodrowski, T. Engel, M. G. Hicks, S. M. Kast, C. Kettner, W. Koch, G. Lanza, A. Link, R. A. Mata, W. E. Nagel, A. Porzel, N. Schlörer, T. Schulze, H.-G. Weinig, W. Wenzel, L. A. Wessjohann, S. Wulle, *Riogrande Odontol.* **2020**, *6*, e55852.

[76] H. Junkes, M. Krieger, P. Oppermann, R. Schlögl, A. Trunschke, H. Weber, FAIRmat - a consortium of the German research-data infrastructure (NFDI), JACoW Publishing, Geneva, Switzerland, **2022**, https://doi.org/10.18429/JACOW-ICALEPCS2021-WEBL01.

**Rasmus Schröder** studied physics and biology in Heidelberg (Germany) and at Trinity College Dublin (Ireland). He received his Ph.D. in theoretical physics, working with H.-G. Dosch at Heidelberg University on the scattering of pseudoscalar mesons. As a postdoc in the lab of Kenneth C. Holmes at the Max Planck Institute for Medical Research, he changed focus and started to elucidate functional structures of the molecular motor myosin interacting with actin. During that time his interests in structural biology and—in particular—electron microscopy evolved. When studying image formation together with the Nobel laureate Joachim Frank, his creativity in technology development matured, leading to his move to the Max Planck Institute for Biophysics, to work on electron microscope development. He then joined Heidelberg University in 2008 as head of cryo electron microscopy. In addition to structural biology, for several years his focus has expanded to include the increasingly important large volume electron microscopy techniques.

**Christof Wöll** received his Ph.D. in 1987 from the Max Planck Institute of Dynamics and Self-Organization in Göttingen after his studies of physics at the University of Göttingen. After a postdoctoral time (1988 to 1989) at the IBM research laboratories, San Jose, USA, he accepted a position equivalent to assistant professor at the Institute of Applied Physical Chemistry, University of Heidelberg. After his habilitation he took over the chair for physical chemistry at the University of Bochum (until 2009). Since 2009 he is the director of the Institute of Functional Interfaces (IFG) at the Karlsruhe Institute of Technology. His work focuses on the development of organic thin films in general and especially on surface-anchored organometallic scaffold compounds (SURMOFs). He is a member of the German National Academy of Sciences, Leopoldina, and holds an honorary Ph.D. degree awarded from the University of Southern Denmark.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

**Stefan Bräse** studied in Göttingen (Germany), Bangor (UK), and Marseille (France), and received his Ph.D. in 1995 with Armin de Meijere in Göttingen. After postdoctoral research at Uppsala University and the Scripps Research Institute, he began his independent research career at the RWTH Aachen in 1997. In 2001, he finished his habilitation and moved to the University of Bonn as a professor of organic chemistry. Since 2003, he has been a professor at the Institute of Organic Chemistry, Karlsruhe Institute of Technology (KIT), and since 2012, also director of the Institute of Biological and Chemical Systems—Functional Molecular Systems (IBCS-FMS) at the KIT. His research interests include chiral ligands and catalyst design for asymmetric processes, molecular engineering of functional synthetic materials, and digitization in chemistry.