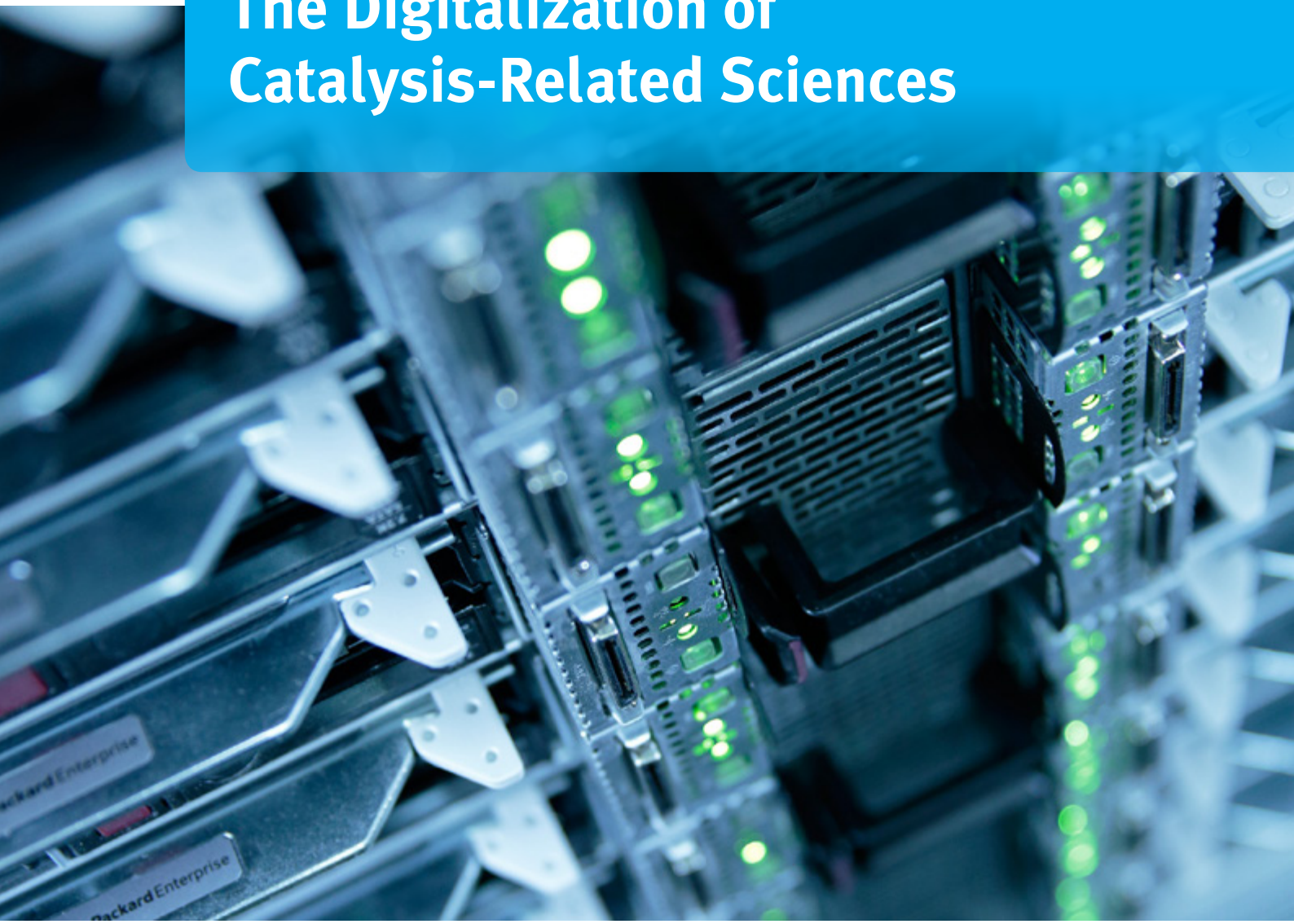
DECHEMA GDCh DBG DGMK VDI

WHITEPAPER

The Digitalization of Catalysis-Related Sciences



GECATS
GERMAN CATALYSIS SOCIETY

Cover Photo: BASF´s Supercomputer QURIOSITY – is an enabler in Digitalization at BASF allowing the calculation of complex model systems in chemistry to speed up development times and reveal hidden correlations, Copyright BASF

TABLE OF CONTENTS

1. Introduction and Motivation

Catalysis as an interdisciplinary technology field is of strategic socio-economic importance. More than 95% of all chemical products are nowadays produced in at least one step using a catalyst which shows the high importance of catalysis for all value chains in the chemical industry. Catalysis plays furthermore a key role in refineries to satisfy the demand for energy, for efficient use of raw materials and for protecting the environment and climate by removal of toxic pollutants. The use of renewable energies and ways of chemical energy storage like addressed in electrochemical water splitting, power-to-chemicals processes, fuel cell technology and conversion of biomass are playing an increasingly relevant role – in all these mentioned fields catalysis is and will be a key technology and enabling element in present and future energy systems. Especially in new and upcoming fields, innovative technologies relying on economic and sustainable catalytic processes are required. In this paper GeCatS addresses the importance of the transformation of the current scientific catalysis scene into a digital catalysis scene.

1.1 Historical and Present Situation

For more than 150 years, Germany has had a leading position in the field of catalysis research and German enterprises efficiently also today exploit value chains in the field of catalysis and related technical field like reaction and process engineering. Many exciting developments and a large number of innovations have been fostered at the interface between the national academic catalysis scene and german enterprises in the realization of new

Executive Summary of the Core Theses discussed in this Paper

Catalysis as an interdisciplinary technology field is of strategic socio-economic importance and will be a core technology enabling mankind to solve pressing challenges concerning climate change, supply of sustainable energy and supply of sustainable materials at the same time. To tackle the mentioned challenges within reasonable time, a step change is required, particularly in improving the development workflows in catalysis.

The core challenge in improving the general understanding and the development workflows in catalysis is building a bridge between theory and experiment. This unification can only be successfully performed applying the principles of digitalization consequently in the field catalysis and transformation of the field into “digital catalysis”.

The vital elements identified in digital catalysis are the unification of data formats and the creation of information architectures that allow storage, exchange and analysis of data utilizing the latest state of the art analysis approaches especially tools based on artificial intelligence.

Digital catalysis requires research-data infrastructure that is user-driven and non-bureaucratic. This will help in accepting digital catalysis on a broad basis and foster participation. In this way, the initiative aims towards a data sharing culture. This is a cultural shift in the world of catalysis research.

The consequent implementation of the FAIR principles in catalysis sciences will lead to large amounts of data sets useful for the community of researchers in the field of catalysis. GeCatS targets at setting the stage for allowing the harvest of the maximum value from these data by initiating their storage in a national data repository for catalysis on a sustainable basis and allowing access on a community basis.

The initiative of digital catalysis will not only lead to a new understanding and a new level in predictivity; it will also help to educate the generation of tomorrow and increase interdisciplinarity between mathematical and theoretical sciences and experimental chemistry, chemical engineering and materials science. The initiative must therefore also be based on consensus and active participation of all parties throughout academia, national research institutions and industry.

and innovative technologies. In this context, innovation is necessary in terms of both the catalyst and the related process technologies to secure the fate of Germany as a competitive industrial location. Due to its high relevance catalysis research in its breadth of the fields of heterogeneous, homogeneous, bio-, electro- or photo-catalysis is known for the fact that the research results are in general associated with great economic potentials. An often-cited famous example of high relevance is the Haber-Bosch ammonia process with a catalyst invented by Fritz Haber, further developed by Alwin Mittasch and a technically realized process by Carl Bosch at the beginning of the last century. Without this process, it would no longer be possible to feed today's world population.

The above-mentioned example also illustrates the complexity of catalysis: A catalyst alone without considering reaction engineering and process conditions is insufficient. This statement is true for all fields of catalysis: heterogeneously or homogeneously catalyzed, bio-, electro- or photocatalytic technologies. Hence, improvement in technical, physical and synthetic chemistry as well as in the field of engineering sciences has led to boosts in catalytic science and technology. This can also be seen in other ground-breaking and pioneering works: Surface science and microscopes with atomic scale resolution have provided a deeper understanding of molecular processes on solid catalysts, in situ and operando characterization has uncovered the dynamics of catalysts under process conditions and high throughput-screening has become a tool for the discovery and efficient improvement of catalysts in all fields of catalysis. Innovative new approaches in reaction engineering have brought new generations of catalysts into commercialization and realization of socio-economic revenues.

Developments in interfacial catalysis (heterogenous, electro- and photo-catalysis), are paralleled by impressive evolutions in molecular catalysis (homogenous and organo-catalysis) as well as in bio-catalysis (enzyme and whole-cell catalysis). The advancements of analytical techniques (spectroscopy and X-ray structure analysis), of synthetic procedures and the advent of computational spectroscopy and molecular electronic structure calculations have enabled the efficient construction of catalytic cycles. These identify critical reagents, coordination geometries and requirements for ligands and metals to effectively steer a molecule through the desired conversion. A resulting systemic approach within molecular catalysis has led to an enormous boost in scientific productivity and to industrial applications ranging from fine chemicals to commodities and bulk chemicals.

1.2 The Evolution of Theory

In the last decade, theoretical approaches in all fields of catalysis have been tremendously improved. The increase of computing power laid the foundations in performing extensive calculations necessary at fast pace. This technical improvement can be valorized, and results can be harvested by a massive evolution of computational strategies, the use of artificial intelligence and implementations of physical concepts [DOI: 10.1038/s41929-018-0188-0ID; DOI: 10.1039/c8cs00398j]. Multi-scale modeling of catalytic processes including micro- and macro-kinetics is today possible [DOI: 10.1039/C7CP07777G]. Building these methods into strategies approaching real-world catalytic challenges as indicated above is however still a challenge to be met. An example is the use of Brønstedt-Evans-Polanyi relationships replacing chemical intuition through computational material science. This approach should allow predictions of performance boundaries beyond thermodynamic limits and of new and alternative material candidates in homogeneous, heterogeneous, bio- and electrocatalysis. Increased computing power forms the basis to perform simulations over several length scales allowing the prediction of practical reactor behavior using microkinetic models. Both approaches in combination allow predicting the performance of a novel catalytic material in a given reaction and in different reactor setups.

1.3 Elements of Realization

Theoreticians in the field of catalysis and materials science have also brought forward an exercise of tremendous importance. Via two vital elements, namely the **unification of data formats** and **the creation of an information architecture that allows storage and exchange of data** big national and international initiatives have been able to create a new momentum in the field of theory and modeling. Notable initiatives that have embarked are the European NOMAD initiative, the American "Materials Genome Project" as well as the "Catalysis-Hub" initiative - all of these share a role-model character.

The two above mentioned elements developed in the field of theory alone do, however, not generate the desired development workflow. Both, for the creation of computational models and for the exploitation of predicted compounds, materials, reactors and process conditions it is essential to build a bridge between theory and experiment. This bridge exists today mainly as a concept without practical viability on a broad and unifying basis. The

deficit is coded in the term “catalysis gaps”. Making the bridge viable is a longstanding desire unaccomplished so far. The conceptual and technical possibilities that become available with the digitalization of science provide now the means to realize this desire. The necessary steps will be outlined in this document. They concern the organization of experimental knowledge. They further require standardization of the vast space of experimental possibilities. They finally require procedures of material creation and verification that do not yet exist in catalysis as in other areas of functional solid material science or in molecular sciences.

1.4 New Potential

The aspiration in the field of “digital catalysis science” is to redesign the workflow for creating optimized or novel catalytic processes. It is evident that to allow for successful implementation, new data analysis tools based on artificial intelligence will play a major role in this context as data of complex structure coming from different fields have to be analyzed. A shift is needed towards “in-silico” prediction of combinations between reactions, catalysts and reactors that form a functional unity fostering the breakdown of a silo-attitude often practiced between disciplines in catalysis. This shift in workflow is not only essential to economize further process evolution but to generate novel solutions in critical areas of chemistry where no success was found with the hitherto applied development concepts. The chemical conversion of small molecules for energy applications (CCU and storage) in central and decentralized application scenarios is a typical area where present scientific approaches can only in parts deliver viable solutions. Other areas are stoichiometric and wasteful synthesis pathways for more complex molecules including the avoidance of harmful intermediates and reactants.

Considering the socio-economic impact of catalysis, it would be highly desirable to expand the path of unification in catalysis of theoretical and experimental approaches and to find a suitable way on how to digitalize the entire national catalysis scene. GeCatS expects that this will lead to another boost of academic and applied catalysis by information-driven design through unified data formats and commonly used information infrastructures.

1.5 Digital Catalysis Science

Hence, digital catalysis science is more than just new software, improved data storage, improved theoretical approaches or digital availability of results. Instead this initiative aims to redesign a science and technology critical to our culture and aims at making use of these new infrastructure possibilities and exploit them for describing the complex nature of catalysts in the reaction environment of a catalytic process. Digital catalysis science will open new opportunities by providing:

- opportunities for comparison and analysis of data of different origin (experimental, calculations) and different disciplines in catalysis on a digital basis
- creation of stronger ties between theory and experiment for improved prediction
- embracing of complexity and thus providing a new understanding with improved predictive tools

In order to deliver these opportunities digital catalysis requires two fundamental technical backbones:

- digital archives for storage of catalysis relevant information in common formats on all length and time scales
- cooperative digital workspaces for data exchange and evaluation of data relating to catalysis

Catalytic technologies and catalysis science in the fields of experimental catalysis and theoretical approaches to catalysis can thus strongly benefit from digitalization. The synergy between theory and simulation with experimental sciences is expected to display step change behavior if fostered adequately. Of vital importance for the success is the opportunity of a new interaction of expertise in catalysis in experimental as well as theoretical approaches. In the light of pressing global and national societal challenges to which catalysis is the key to contribute solutions, it is from the point of view of GeCatS evident, that a digital transformation of the catalysis community must become a national mission. GeCatS believes that a national initiative for the digitalization of catalysis should also be open for contributors and contributions from the European and international catalysis community, one a certain stage of maturity has been reached. This mission is in accord and alignment with the European vision and the targets of the European Open Science Cloud (EOSC) [https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud]. Hence, GeCatS has started an initiative to create digital catalysis science.

2. Vision

GeCatS' vision of the "digitalization in catalysis initiative" is a next level of catalysis research that allows not only to compile and mine data for various catalysts but also to describe and understand the complexity in catalysis and predict improved catalysts, rational reactor designs and optimized process conditions. Hence, GeCatS is convinced that a digitalized scientific approach in catalysis and a transition into a digital catalysis community will make a major contribution to the progress in the field of catalysis. The major values that a digitalized catalysis scene can bring forward will be described in the following sections. These values are generic and rest in science. They are primarily not technical or economic values. They are, however, indispensable elements of better technologies and improved economics. Hence all stakeholders in catalysis, irrespective whether academic or industrial, will benefit from digital catalysis science. GeCatS strongly believes that the community of scientists active in the field of catalysis must be in the lead to create and define the transition and the boundary conditions of digitalizing catalysis.

2.1 Unification of the View on Catalysis

The central vision of digital catalysis science is a unified view on catalysis in all dimensions. On the process side the unified description between catalyst, reactor and reaction conditions must be obtained for rational planning and a causal understanding of a catalytic process. The far-reaching consequence is that in the analysis as well as in simulations the traditional separation between "chemical elementary steps" and "engineering transport phenomena" must be superseded. The reason for that lies in the aspiration of digital catalysis science to analyze and predict performance processes and not to confine to model conditions. As essential and necessary model systems and their reactions are for establishing insight into elementary processes as clear it is that model conditions cannot be extrapolated to performance conditions. The influence of mass and energy transport on one side and the creation of materials far off the thermodynamic equilibrium situation during operation on the other side both create this "gap" that should be bridged by a transition into digital catalysis science.

On the catalyst side the well-entrenched separation between disciplines must be removed. Catalysis is a universal concept of chemical kinetics dealing with the modification of reaction rates by affected by energy barriers of elementary reaction steps. This concept is valid for all classes of catalyst materials and for all forms of energy sources (possibly not equivalent for all energy sources). Digital

Executive Summary Core Statements "Vision"

GeCatS believes that a digitalized scientific approach in catalysis and a transition into a digital catalysis community will make a major contribution to the progress in the field of catalysis.
The central vision of digital catalysis science is a unified view on catalysis.

Digital catalysis science wishes to learn from experimental performance observations and wishes to predict the performance of a (novel) catalysts, therefore it must include all multi-scale effects which also link strongly into industrial applications.

Core topics important for the success of digital catalysis are common information architectures and adequate and community accepted data formats that need to be established. They will allow interactive exchange of groups and individuals, sharing and analysis of data, bridging exchange of data between different disciplines across the whole of the catalysis (model catalysis, interfacial catalysis, molecular catalysis, bio-catalysis and engineering disciplines) community.

The technology field of Artificial Intelligence is a substantial strategic technology element and a core enabler which guarantees success in the transformation process towards digital catalysis due to enhanced analysis capabilities of complex datasets.

2. vision

catalysis science should make use of all insight generated for a given reaction from all types of processes and materials. The widely used prejudice that molecular catalysis is unpractical on a technical scale is as questionable as is the prejudice that only in the field of thermal molecular catalysis reaction sequences are understood and all other forms of catalysis do this to a much lesser extent. In a future research landscape with digital catalysis science at its core, strong links between of the research areas of model catalysis, interfacial catalysis (catalysis on surfaces), molecular catalysis and bio-related catalysis with the goal of generating a unifying view for the benefit of a joint understanding of a given reaction. The link between these traditional independent areas is a common underlying rigorous physical science. The medium required for this link is data concerning structure and function in the respective area. Figure 1 illustrates the interrelation of the different research fields that need unification for a complete picture of a catalytic process.

It is only through the potential of digital science and theoretical tools that the enormously diverse data from different sources and disciplines can be collected and represented in a form allowing to grasp the complete picture of a reaction with its dimensions in space time and chemical materials involved. It is a core benefit of digital catalysis science that the hitherto necessary simplifications of conceptual complexity can now be safely removed, and a still rigorous description remains a realistic task. An all-embracing sharing of data (including metadata) will advance catalysis science and engineering because shared data about studies do not need to be repeated, and studies performed for a certain purpose can be used for addressing a different purpose – so cooperative learning of the community is fostered.

In future, traditional experimentalists and theoreticians need to jointly tackle the understanding of catalysed reactions and the mode of action of the catalysts employed and thus must develop a common language. At present, apart from single point contacts of the communities, no general synergies over a broad front can be observed. A step change in the field of catalysis can thus only be expected if an approach that allows communication and sharing of results on eyesight level is achieved. From a GeCatS view, we thus envision information architectures to share and work on data derived from experiments and theory/modelling on a basis of a digital community. This requires the convergence of the experimental and theoretical communities in terms of data definitions and language as well as in terms of methods, tools and procedures. Standardization of minimum informational requirements and quality control of experiments will be needed to generate “clean data” for digital processes. It should be noted here that such quality control procedures suffer currently from lim-

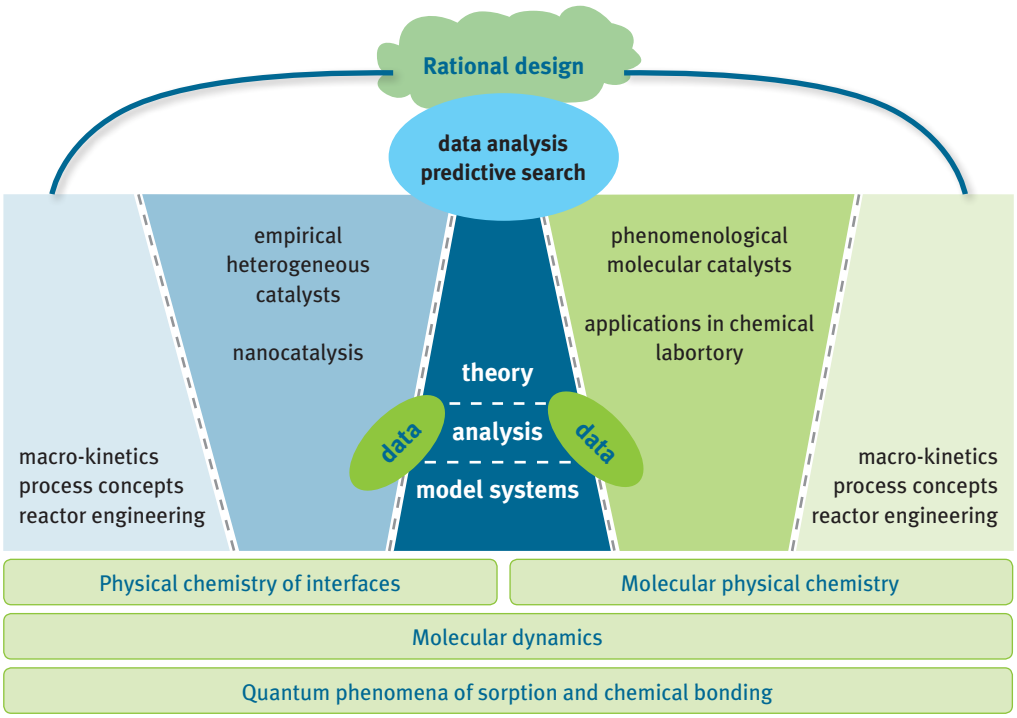


Figure 1: “Map” of catalysis science and their interrelations. The dashed lines describe paradigmatic boundaries that need to be overcome for a complete picture of catalysis.

ited popularity in the field of catalysis science today, and initiatives for cultural change in academia have had limited impact so far [DOI: 10.1021/acscatal.6b00183]. Based on these facts it becomes clear that an enormous challenge is associated in including past knowledge published in the open and patent literature as input into emerging digital catalysis science. One suggestion to be evaluated must be to designate data sets both in theory and in experiments with quality assurance descriptors (like the "reliability star system" in structural data bases). This allows incomplete and unqualified data from the past to be included in, for example, machine-learning procedures for constructing descriptors allowing rational searches in material libraries. In the future libraries of high-quality data from experiment and theory will be needed to validate such approaches and eventually to converge to reliable descriptors. One promise of digital catalysis science is certainly that powerful algorithms may interpolate over incomplete data sets and thus allow a certain tolerance against partly incomplete (noisy, disturbed or incorrect) data within a library. Such options are a great benefit of digital catalysis science but need massive efforts before knowing about the robustness of this vision.

2.2 The Active Site

Active sites are the central element in all catalyst systems. In molecular catalysis and in certain cases in the field of bio-catalysis it is possible to accurately determine the composition and structure of active sites. In interfacial catalysts they represent a fraction of the surface (reactive surface area) and can occur in multiple geometries; knowing the exact nature of active sites is thus a challenge that can greatly benefit from computer experiments simulating active sites.

A challenge for all catalytic materials lies in the fact that active sites are non-equilibrium structures and can only catalyze a reaction because they represent "high energy" sites. Physical models of such high-energy sites are unlikely to exist, and existing structural models (crystals) of active sites are not highly active. Here computational modelling can suggest possible structures of active sites and their relation to the local environment from which they form under reaction conditions. Evidently computational modelling here can benefit massively from insights generated via operando analysis. If this can be achieved for many systems one may obtain insight into the formation principles of active sites and enable their frequent occurrence by adapting the local solid or liquid environments.

At present, active sites are described by a minimum of descriptors indicated in the Figure 2.

Stepped solid surfaces or metal atoms with a first sphere of coordinating ligands are typical examples. The insight gained into active structures from using in-situ and operando analytical techniques revealed that such a static picture of the active site is a boundary condition at vanishing conversion and in most cases not sufficient. The concept of active sites is even more complicated by the fact that there is a distribution of active structures and hence a

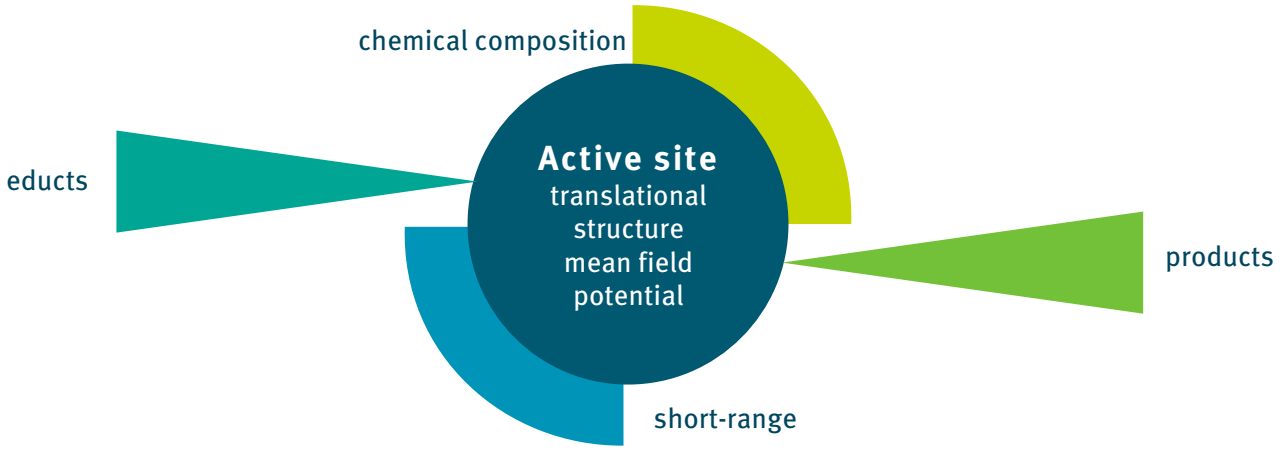


Figure 2: Descriptors of active sites. Chemical composition and the first coordination sphere (rarely also the second sphere) are used for defining the active site. There is a tendency to even remove this coordination and deal with active sites in terms of single atoms. It is further recognized that active sites are subject in their structure to the coordination of reactants and products, which influence with their average chemical potential the average structure of the site.

2. Vision

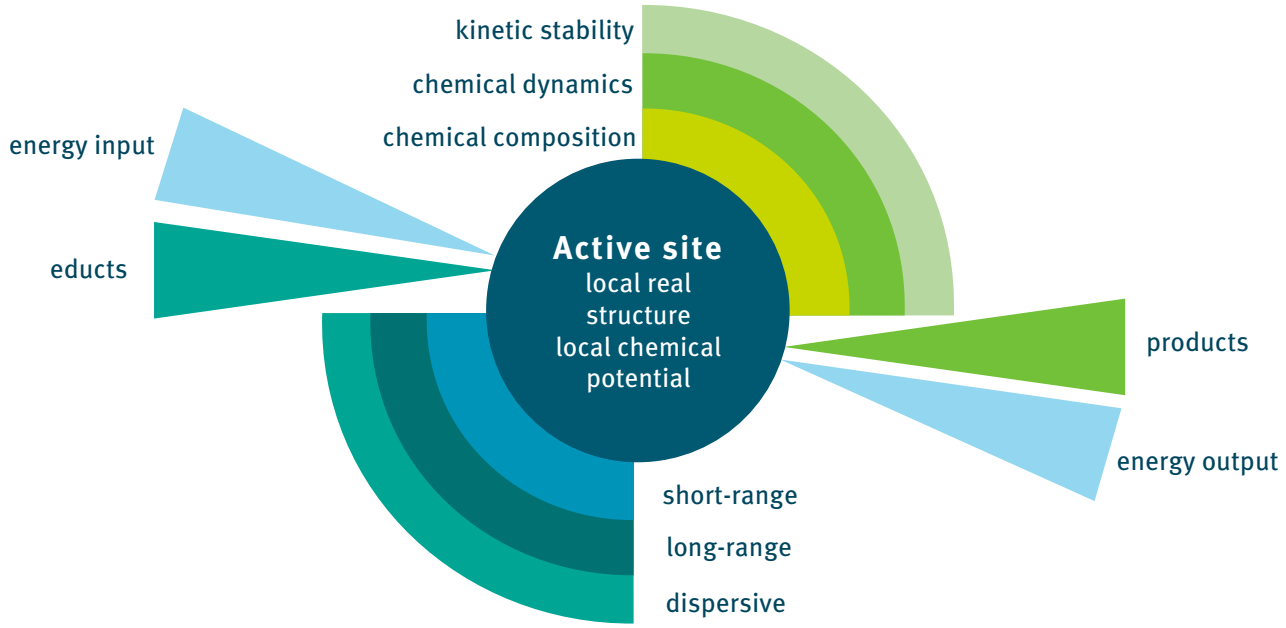


Figure 3: Descriptors needed to characterize active sites at performance conditions where they become dynamic and cannot be approximated any more with a static picture. The time-dependent variations of active sites are largely controlled by multiple weak interactions affecting the local vibrational dynamics of the active structure. Functional concepts of enzymes, solvent effects in molecular catalysts and co-adsorption phenomena at interfacial catalysts are typical representations.

distribution of kinetic parameters. Under reaction conditions of high severity like i.e. high degrees of conversion in industrially relevant reaction condition corridors, the characteristics of the active sites become highly time dependent in relation to the overall molecular dynamics at a catalytic site. If reactions with macroscopic selectivity to one desired product are under study (and this is the typical case) then multiple pathways of a reaction through the energy landscape can exist and the design of a maximum performing catalyst involves much more control than finding the minimum barrier path. In such a situation the description of active sites needs substantial addition as indicated in Figure 3.

These descriptors act on the structure of active sites on vastly different time scales. For that reason, it is inadequate to describe the action of a catalytic site only at the time scale of electronic exchange processes – still this appears adequate if the function of the site lies in breaking or creating chemical bonds. In any real situation the site is not alone only with its reacting molecule and this conceptually complicates the situation substantially. Through this coupling of time scales also dimensional scales are important in defining the distribution of active sites. Transport phenomena on multiple length scales determine the local chemical potential of an active site. If digital catalysis science wishes to learn from experimental observations or wishes to predict the performance of a (novel) catalyst as molecular entity, biological entity or material, then it must include all these multi-scale effects. The experimental way out to perform single molecule transformations and to eliminate the system inherent complexity is not recommended in the present context as the relevance of single molecule trajectory for the description of the production of even a small amount of product is limited.

2.3 Mastering Complexity

The concept of working active sites unravels levels of complexity that cannot be handled by conventional scientific approaches. This has forced catalysis science in the past to either make simplifications (the static active site) or ignore the rational design (combinatorial design). Most of the conceptual insight that is available today came from these simplified approaches. Yet it is necessary to combine them and omit the conceptual limits that have created the “scene gaps” if the challenge of the design of catalytic processes is to be met.

GeCatS expects that a digital approach in catalysis allows mastering the inherent complexity much better. This can be done in several ways. First computational modelling can cope with the multi-scale challenge and explicitly describe active sites in action. At present this is still limited to simple reactions. There are several reasons for this

unfortunate situation: lack in communication between theory and experiment is one of them, but on the other hand insurmountable limitations in adequate yet to be developed methods in theoretical chemistry of are another limiting factor. The communication gap further hides the fact that in the experimental domain many observations of a given system are still missing that would be needed for a rigorous understanding. Paradigmatic simplifications (ignorance/simplification of reaction networks of all reactants) and a still small number of true operando system observations (meaningful spectro-microscopy at simultaneously recorded kinetic behavior) reduce the opportunity for applying the concept of digital catalysts science by analyzing existing systems through the machine learning approach. This rational knowledge-based approach is illustrated in the central pathway of Figure 1.

An orthogonal approach to the explicit incorporation of a complete set of phenomena on all scales of time and space is to use computational modelling to make a sensitivity analysis and to select those processes that can be included with a simplified descriptor against those requiring an explicit description. With the advent of Artificial Intelligence, it may even be conceivable to extrapolate over some of the phenomena and reduce the precision of the digital prediction for the sake of a much-reduced complexity per case to be considered. This ability of theory to study catalysis at various levels of complexity and thus choose between precision and width of description was already instrumental in the development of theoretical catalysis science. Historically at the turn of the last century it was clear that the full description of a catalytic reaction would require enormous efforts in the mathematically correct description due to inherent complexity and the explicit bridging of multiple scales in time and space. This “precision” path led to the development of all the computational techniques that are available today ready for digital catalysts science. Orthogonal to this the “predictive” path was developed. Here the description of a chemical reaction is reduced to its most important energetic components and the definition of one elementary step as kinetic bottleneck. By casting this into a static model of active sites (Figure 2) it becomes possible to search through a great number of potential catalytic materials and arrive at a trend that describes families of potential and often novel candidates and families of systems that are clearly unsuitable. Narrowing in such a way the search space for experimental approaches is a most valuable and highly effective support of catalysis science. This approach can also be performed for reactions that cannot be catalyzed today at all. Its results allow for gaining deep insight into the origin of this deficit. Another mode of operation would be the *in-silico* differentiation between different process pathways to a given molecular structure. A coarse road map with the alternative options would allow estimating the feasibility and chemical challenges to be expected before having performed a single experiment.

These flavors of digital catalysts science can be performed with a much lower barrier of introduction. They are extremely useful in practical catalysis science as they allow focusing the resource-intensive experimental efforts by reducing complexity in a guided rather than in an intuitive way. The range of applications is wide as illustrated in Figure 1, since both interfacial and molecular systems can benefit.

2.4 Computer-Aided Experimentation

The ultimate vision of digital catalysis science is to construct for a given target molecule the energy landscape for any envisioned synthesis route and to suggest catalytic systems consisting each of catalytic materials, a reactor configuration and a set of reaction conditions. Experiments would have to verify the suggestions and provide the evidence for decisions that need much augmented information from outside catalysis science.

This mode of operation is still far away from the position where the idea of digital catalysis science is today. Critical steps towards this vision are, however, the simultaneous co-operation of theory and experiment based upon mutual complete information of results through structures and interpretable data. Theory can then suggest experiments either in the analytical or in the synthetic realm from which a maximum of new information can be extracted. These suggestions from theory do not necessarily lead to successful catalyst and reaction design via a single “in-silico”-shot, but they can help through the gathering of experimental data to improve the models and to close the development cycle faster. A situation in a digital catalysis community with a common information architecture and a series of common data exchange formats is a prerequisite for changing the scientific approach and enable a better guided experimental science through augmentation by theory modeling. A core aspect of digital catalysis is also to foster new methods and alternative approaches in theory and modelling.

One particular area where digital catalysis science can make a strong impact is computational spectroscopy.

Today the interpretation of spectra is largely based upon comparison to seemingly known systems (fingerprinting). This is proven to be a quite dangerous approach creating a vast number of wrong interpretations. Both false reference systems and multiple methodical artifacts merge into a dangerous alliance leading to many wrong analytical statements in the existing literature. If all these would enter digital catalysis science, the resulting prediction may be quite unreliable. To create a library of computationally assigned spectra and to create tools for adequate data analysis is a low-hanging fruit improving the experimental input into digital catalysis science and helping in identifying false entries in the process of machine learning from pre-existing data.

GeCatS has identified that as core topics common information architectures and adequate and community accepted data formats need to be established. They will allow interactive exchange of groups and individuals, sharing and analysis of data, bridging exchange of data between different disciplines across the whole of catalysis (model catalysis, interfacial catalysis and molecular catalysis) community. Data exchange is needed especially between experimental groups and groups being active in theory and modelling. Such information infrastructure will be far more valuable than any of the current collections in the open literature, as it will also serve as a "lessons learned" data base, realizing that approaches that have not led to improved catalysts systems are still valuable data to solving a catalytic challenge. Beyond that such an information architecture will be a repository of corporate knowledge in catalysis, data for case-based reasoning, independent assessment, review options, and services. Finally, this database containing information on catalysts, their specific properties, their behavior under certain reaction condition and the like can also serve as tool for temporal trend analysis in the field of catalysis.

GeCatS is convinced that common information architectures and adequate and community accepted data formats will allow a significantly more efficient way of tackling the general questions of catalysis including its consequences for globally pressing topics such as energy generation, energy storage and finally climate and resource protection. This will eventually lead to new methodologies going far beyond the sole use of such an infrastructure, and the aim is that this also includes the nature of information and data quality as well as the different ways of their generation.

An additional opportunity of such common information architectures and adequate and community accepted data formats are that the comparability of experimental data will lead to the possibility of referencing including reference tools and reference methods – a long standing topic of debate in the field of catalysis.

2.5 Artificial Intelligence Concepts as Key Enabling Elements in Digital Catalysis

In the context of this chapter "artificial intelligence (AI)" is used as umbrella term of computational methods that "learn from experience". This includes the numerous methods of machine learning as well as methods that originate from signal processing (e.g. compressed sensing), data mining, and more. An integrated data view for data stemming from experimental and theoretical work is one of the core preconditions that must be achieved with the digital catalysis approach. As illustrated above, the complexity of the phenomenon catalysis itself, the intricacy of the various actuating processes, and the complexity of data sets resulting from experimental work or modelling can more efficiently be tackled using AI Concepts. GeCatS believes that AI Concepts are critical strategic technology elements which guarantee success in the transformation process towards digital catalysis. The revolutionary advancements of AI have disrupted the field of data analysis and helped in science to bring forward new insights and to establish new ways of interaction in a range of business domains. In digital catalysis the sheer amount of data that is expected once user groups use an efficient data infrastructure, will be a great opportunity to work with the technology and develop applications that are most suited for the field. Apart from the opportunity of analysing large amounts of data, it is also expected that AI will help in creating bridges between the different disciplines in catalysis and therefore also be a fostering element in collaboration of the different disciplines through unified data analysis. AI may identify correlations between different data sets or anomalies that are invisible when studying small data. Therefore, it is also expected that the use of AI will help in creating a more systematic linkage of data from (multi scale) computations and experiments.

Another strong argument for embarking simultaneously on the development of methods in AI specific to catalysis simultaneously with the rollout of unified data formats and an integrated information architecture, is that a maximum advancement is expected in the co-development and application of new methods in AI, the development and use of unified data formats and the use of the integral information architecture. Computational data from theoreti-

cal and simulation approaches and experimental data can and should be used for analysis through artificial intelligence tools. As the data are typically not really “big” compared to other data sources where AI is applied, domain specific AI approaches need to be developed. Experiences from other fields in science where AI has been successfully applied show, that algorithms in the field of AI which are specifically adapted or developed have helped in tackling “tough nuts” are also expected to reveal so far hidden relations in catalysis which help advance the field in areas where progress is needed. Catalysis is controlled by many complex and intricate processes. It is assumed that the desire to “understand” catalysis in its depth at real-life conditions cannot be reached in the first place, even with AI as helpful toolset. However, AI will help will provide a reliable description and it will help in the process of identifying suitable descriptors that allow to derive a reliable description on a cognitive basis [C. Draxl and M. Scheffler, Big-Data-Driven Materials Science and its FAIR Data Infrastructure. Plenary Chapter in Handbook of Materials Modeling (eds. S. Yip and W. Andreoni), Springer (2019).].

The final goal that appears viable through the application of methods based on AI, is to come from a mode of work that relies on “descriptive analytics” to modes of work that relate to “predictive analytics” and for a range of application examples also to modes of “prescriptive analytics”. Especially the two fields of “predictive” and “prescriptive” analytics are up to this point in time not advanced in catalysis – but would represent highly desirable modes in the light of pressing societal challenges.

3. Mission

GeCatS vision for a digital future of catalysis requires substantial modifications of the paradigm how catalysis science is performed today. This s likely only to occur if four prerequisites will be met:

1. The transformation needs to be sustainable. This means that an initial project creating the technical and conceptual environment needs continuation beyond the volatility of the present project structure. One part of an envisaged research project for the creation of digital catalysis science will be to define a roadmap to sustainability. Continuous funding and accessibility of the digital resources are prerequisites for the broad acceptance of the approach.

2. The infrastructure of digital catalysis must be open and allow for diverse approaches despite the necessary standardization of the data structures. Any rigid and privately-owned element in the infrastructure is to be avoided. Multiple options for data entries and open source software tools are essential for the broad and continuous acceptance of the whole concept.

3. There must be revenue for those who participate in digital catalysis and share their data or accept standardized experimental procedures. This revenue can only be the unlimited and toll-free access to the results collectively generated by the digital community and the unambiguous labelling of the origin and authorship of data-sets in citable format. Scientific work that creates high quality data must be rewarded, beyond the referencing of publications and their metric such as h-factors. The creation of high-quality data must be coupled to appropriate funding.

4. Such a working structure requires regulations for data ownership, IP protection, data security and a scientific reward for giving away own research data. The issues of plagiarism (self-plagiarism) and verification of results need also attention.

Consequently, it is the natural business of a society like GeCatS as the platform in Germany for catalysis science to take the lead and structure the discussion process in the community as well as to serve as competent expert to the funding agencies and other stakeholders (industry, publishers, instrument vendors, software developers). It is most relevant for the trust of the community in such a transformation that these burning questions that are not in line with the existing scientific practice are tackled with

Executive Summary Core Statements “Mission”

The digital transformation of catalysis needs to be sustainable, therefore the infrastructure of digital catalysis must be open and allow for diverse approaches despite the necessary standardization of the data structures.

Unlimited and toll-free access to the results collectively generated by the digital community must be guaranteed. Regulations for data ownership, IP protection, data security and a scientific reward for making own research data generally available to a digital community are required.

The FAIR concept for the treatment of data must respected on all levels as major design criterion. Five pillars of data formats are seen as fundamental for digitalization in catalysis.

An initiative for digital catalysis science should consider that massive bottlenecks in documenting chemical complexity must be overcome as the proven realization of novel systems with high performance developed by digital catalysis in lighthouse projects will convince all stakeholders about the usefulness of the new science paradigm.

The development of automated agents that allow seamless reformatting of user data into the community agreed formats will lead to community acceptance through user-friendly incorporation of such agents into local digital landscapes like electronic laboratory journals.

the same serious efforts as the intrinsic scientific-technical topics.

The digital transformation of the scientific community working in the field of catalysis can only become a successful endeavor, if the transformation approach is regarded as holistic endeavor. A joint interdisciplinary approach based on common data formats is required. Hence, measures must be taken to not only set the technical basis for the digital transformation, but also allowing individuals, organizations and the community to actively participate in the transformation process allowing shared control, shared ownership and setting the stage for an active community in the field of catalysis. The experiences of industry and public organizations in digital transformational endeavors show that the resources, that must be accommodated in human capital on the user-, organization- and community-level must not be underestimated and – even worse – are decisive for the successful implementation and application of investments in resources made on the technology side [Siemens Financial Services | Frühjahr 2018 | Praktische Wege zu Industrie 4.0]. Figure 4 shows an illustration of the different hemispheres that are concerned in the transformational process; all parts of the "digital hemisphere" must be considered to allow for a successful transformation. Exactly because most of these experiences gathered and lessons learned in digital transformation processes have been gathered in industry, the participation of industry is seen also as a key critical component within a successful digital transformation process of catalysis-related sciences via a community approach.

3.1 Technical Prerequisites

The focus on a technical basis should be on the two main enabling elements allowing the interaction of individuals and groups within that community. These two elements are:

» Common information architecture following the requirements set by the community

» Common and accepted data formats

The prime and first elements that serve as a basis for interaction of the community and have platform character are common information architectures. For a digital community in the field of catalysis the meaning of common information architectures even goes beyond the pure technical function as a data-base-backbone, as these architectures also have specifying character in defining the purpose of the digital community. In a digitalized catalysis scene, common information architectures are envisaged as universal platforms, being the vital backbone for storage, collection and organization of data in common and accepted formats. To allow these information architectures to become common workspaces, allowing groups and individuals to work with the data contained in the archives in supporting findability must be addressed by key functions that address aspects of organization, labelling and navigation. Aspects like access rights and the establishing restricted work spaces have to be addressed in an adequate manner, in order to render these architectures also viable as platforms in daily project work.

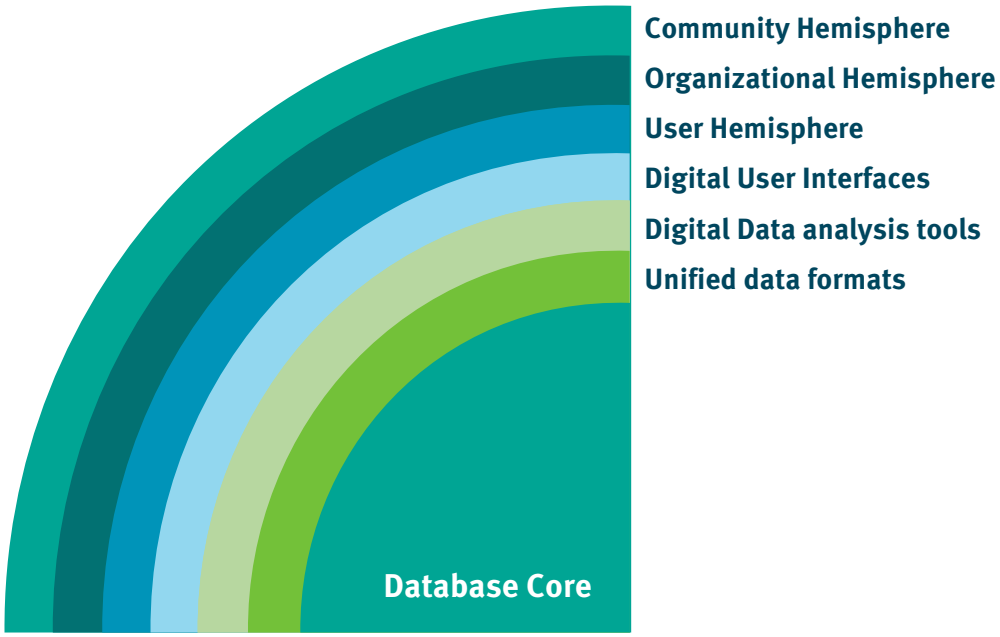


Figure 4: Digital hemisphere model illustrating the sequence of hemispheres involved in digitality.

In addition, the information system should provide capabilities of an interactive shared information architecture allowing accessing, browsing, addition and updating of data.

An integrated data view should allow users to get access to data in an efficient manner via as few queries as possible. With up-to-date cloud-based information architectures efficient structures, mechanisms and functions can be brought in place that do have to rely on data consolidation in a single system but can operate via data-virtualization or data federation processes. Another important requirement regarding the structure and functionality of the information architecture is extensibility with regard to third-party tools; to get the full value out of the information architecture the opportunity to seamlessly link third party tools should not be missed. Of specific value for the community are especially tools which have open-source character and are therefore also open for modifications by individuals within the community.

The second vital technical element on the pathway to a digital catalysis community are common and accepted data formats; common data formats are the enabling elements and define *taxonomy (general principles of classification of entities) / ontology (structure of entities within a domain)* and *content* of sets of data and associated meta-data and must be the result of a process of community consensus. The goal of common data formats is to enable individual users to work in shared information environments. Common and unified data formats are the basis for seamless communication within and in between the different disciplines of catalysis. A unique value arises in the potential of creating bridges between parts of the community working experimentally and other parts working based on theory and modelling. Especially this last-mentioned link between the two different areas in catalysis theory and practice/experiment has an immense value proposition: Data generated in both fields can be compared on objective basis if an alignment in common data formats is achieved. The value proposition behind this connection of the two fields is bi-fold: on the one hand predictions based on calculations from theory and modelling may lead to better designed experiments, on the other hand treatment and analysis of data obtained from experiments by experts from theory, modelling and simulation may result in better understanding and higher predictivity.

3.2 Five Pillars of Catalysis Specific Data Formats

Data formats required for catalysis are unique, largely due to the nature of catalysis as an interface phenomenon including states far from thermodynamic equilibria which can only be described adequately via kinetic phenomena in combination with transport properties. The description of the phenomenon “catalysis” requires therefore data formats that allow a representation of a multidimensional picture of reality. These unique requirements regarding the specificity of the data are mainly reflected by the fact that the catalysis community has found forms of analogue representation which are more or less well documented and respected within the community.

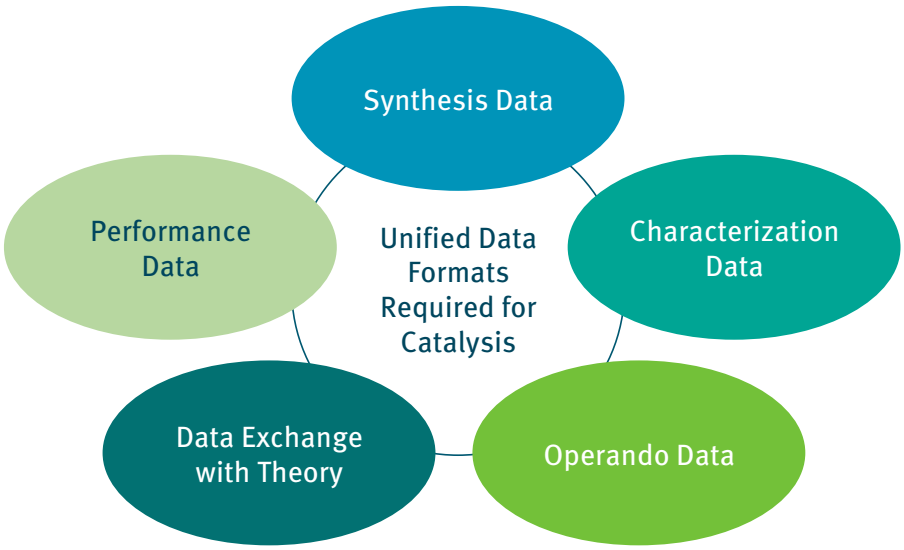


Figure 5: Five pillars of data kinds of formats seen as fundamental for digitalization in catalysis.

From a current perspective five kinds of data formats are foreseen as mainly relevant families of data formats with regards to fundamental functionality of a data repository useful for digital catalysis. In contrast to the present practice all these pillars need to be filled with content for a meaningful description of a catalytic process. And in all these five pillars a suitable level of meta-data (experimental details) are required, a condition that is barely met today. Publications are more directed towards telling narratives about catalytic reactions and the documentation of experimental results and details is likely to be negatively connoted as "over-detailed" as it gives seldom rise to scientific credentials. The technical prerequisites for a complete description of "clean data" do exist. This practice must change to arrive at the necessary content for the five data pillars.

Performance data

Performance data (typically activity, productivity, selectivity and time-dependent behavior more generally described as stability or turn-over-number) relate to a defined stoichiometric composition of chemical compounds including a catalyst and a defined set of physicochemical variables which the catalyst and the reaction mixture have been exposed to. Here the issue arises that many catalysts are not known in their composition during operation and one conveniently refers to the stoichiometry of the pre-catalyst or the assumed molecular structure. For interfacial systems a chemical composition is almost meaningless as they often are phase compositions with critical morphological and chemical interrelations (supported catalysts on non-stoichiometric surfaces of a support phase). The practice of defining catalytic activity as turnover frequency with simply assumed numbers of active sites as reference needs to be complemented with a full report of the true observed performance data. Performance data must be strongly related to the experimental conditions that they were obtained at and the reactor type and specifications used and have therefore a strong link into engineering sciences and must therefore also be aligned with the engineering community.

Synthesis data

Synthesis data are typically associated with catalysts being part of the groups of molecules, compounds and materials. The complexity of synthesis data is typically higher when associated to catalytic processes involving materials e.g. heterogeneous, electro- and photocatalysis; they relate to stoichiometric compositions and a defined set of physicochemical variables. Typical sets of physicochemical variables are applied in a sequence to change the nature of the material to be applied as catalyst. Here an exact description of the workflow with all details is essential as an active catalyst with its high energy sites is always the consequence of a kinetic process leading to an incomplete reaction with respect to a thermodynamically stable system. This holds likewise for molecular systems where the formation of the active phase may well only occur in the reaction system. Synthesis data must also comprise data regarding technically relevant unit operations and sequences of unit operations used and indispensable on laboratory and on industrial scale.

Characterization data

Characterization data relate generally to catalysts as molecular compounds and biological entities or solid materials. The type of data depends completely on the characterization method applied and may therefore contain spectroscopic information, information from scattering, diffraction, microscopy, or other physicochemical methods. The community will have to agree on a minimum of characterization data which is essential for extracting meaningful structure-property-relationships. On the other hand, larger and more diverse sets with additional data are expected to be available for selected systems and for systems which have been originally prepared and characterized for alternative purposes. Here the digital information system will need the flexibility to store additional data besides those declared as mandatory. It is essential not to lose the information connection between additional data and mandatory data.

Operando data

Operando data are time series of characterization data related generally to catalysts being part of the groups of molecules, compounds and materials under conditions under which a catalytic reaction is performed and where the catalytic data are collected simultaneously. The type of data relates on the one hand to the characterization method applied and to performance data obtained. Therefore, the type of data has features of characterization and performance data types.

Formats for data exchange with theory, molecular and multiscale modelling
The exchange of data between groups working in the field of theoretical chemistry and modelling itself and with groups working experimentally is another important pillar of the data formats required for the field of catalysis. Here interfacing and translation is one of the main topics that must be solved.

To design the data formats of the five pillars as functional and sustainable as possible, certain important criteria must be considered. To arrive at an accepted taxonomy and ontology a community driven approach is the preferred vehicle of choice for a consensus-model. This consensus must take place on a horizontal basis of a discussion between the different disciplines in catalysis in accord with engineering sciences – the vertical basis of a discussion between academia and industry is also an important one, especially in the light of industrial experiences and potential advantages and shortcomings of the details of certain approaches. It is important also to consider in this context that the ability to grow the taxonomy/ontology by adding new terminologies, variables or relationships between existing variables plays a large role, as this design criterion allows adjustment processes within the data formats, without compromises on functionality of the information architecture. Another aspect is that the ability to add to a taxonomy will also support new (known or unknown) use cases in the field.

Another important aspect that must be implemented in the principal design criteria of data relating to one of the five pillars, is that all such data-sets must be associable with metadata. Such metadata may relate to aspects of the data which are not represented by the data as such and may be in relation to other sources of data. A typical example for distinct metadata is text-based input: in many cases publications like patents, papers or books. The importance of such metadata cannot be underestimated as these data-sets will allow a context-based interpretation of boundary conditions which are in many cases hard to grasp, once the originator of a data set is not in dialogue with users of the data set. Other aspects that add to the value of metadata are parameters for quality and completeness, links to related data-sets, data-sets relating to validation of data and finally community comments and discussions.

To allow for acceptance in using the unified data formats and conversion of data sets from experiment and theory, a maximized convenience for the user must be achieved. Any additional workload imposed on the user community regarding tedious workflows in data-format-conversion will lead to compromises or total rejection of the strategy of unified data formats by the users. GeCatS is convinced that this aversion can be circumvented by the development of agents which perform the conversion of data from experiment and theory fully automatically into unified data formats, which are then machine readable. Such agents should also be able to perform consistency checks transparent to the user and data-entries into the system should be based on user approval. For groups and individuals being part and participating in a digitalized catalysis scene the use of individually or group-owned electronic lab-books or -journals, data-bases and laboratory information management systems is a natural precondition. Therefore, a seamless link of agents capable of converting formats of data available in the aforementioned systems into formats ready for submission into a national data repository for catalysis seems the most practical way for researchers to deal with the conversion of their data into the community-agreed formats.

3.3 The FAIR Concept

To allow for a maximized convenience regarding data exchange in between different platforms, it is also vital to offer a range of different data conversion tools, which can convert information from one data format into another. This capability will ease the adaption of already existing systems like for example the NIST or the NOMAD databases.

Massive synergies are expected from common and accepted data formats and common information architecture following the requirements set by the community:

- The building of strong ties between groups working in modelling and theory and groups and individuals working experimentally and putting these two groups on eyesight level can induce a complete new scientific dialogue.
- Scientific dialogue is enabled and simplified in between the different disciplines in catalysis, as clearly defined data formats will allow a wholistic view beyond each single field.
- Better link to engineering sciences will be established as data exchange formats can be better defined and lead to clarity in communication.

The FAIR concept for the treatment of data must be respected on all levels as major design criterion. The cataly-

sis community must be convinced that FAIR is an umbrella definition and scope for dealing with data.

In detail the principles are [DOI: 10.1038/sdata.2016.18]:

» Data sets should be **FINDABLE**

» Data sets should be **ACCESSIBLE**

» Data sets should be **INTEROPERABLE**

» Data sets should be **RE-PURPOSABLE** or **RE-USABLE**

3.4 Chemical Prerequisites

Catalysts are functional molecules, biomolecules or solids with their function depending on extrinsic properties of their structure. Ligand instability, defects and chemical dynamics are the vital ingredients that discriminate catalytically active and selective catalysts from inactive and unselective catalysts. Hence, neither their stoichiometry nor their stable analogues (single crystals) are the required forms of catalytic materials. If theory predicts a certain stoichiometry as interesting it needs also to specify in which defective form or molecular environment this should be synthesized. Currently this is far from reality and thus a good deal of experimental creativity is needed to synthesize samples of truly catalytic materials and molecules with the desired electronic structure and with a practical stability. A similar argument holds for formulating and shaping solid materials or for creating reactive liquid environments for molecular catalysts.

Chemistry has not yet provided a standard technology for “rapid prototyping” of catalytic materials. This is a bottleneck also for the technology of high throughput experimentation. The need to be able to synthesize compounds in discontinuous phase spaces and to control the real structure such that homogeneous but reactive systems result, is well recognized, yet little scientific effort was dedicated so far to arrive at a generic solution. It may well be that such a generic solution does not exist and that powder samples as well as thin films besides molecules in various mixed liquid phases may be needed in parallel to serve the needs of testing reactivity.

There are very few places where such synthesis can be executed with the necessary diversity and precision. The verification of the integrity and questions of phase purity pose additional challenges when the post-synthetic activation procedures are concerned that convert the pre-catalysts after synthesis to its active form (mostly in the testing reactor). All this needs to be well documented with stringent taxonomy and ontology as in later stages of verification the scale-up of material synthesis requires exact recipes. Material synthesis is a kinetic process as the catalytic reaction intended, and it needs the same careful control of reaction conditions.

Another challenge consists in the transfer of knowledge from the microscopic, laboratory scale into the macroscopic, real world scale, because in technical reactors behaviour of a reaction mixture may significantly differ from any model system, change its state and structure in operando, and interacts with mass and heat transfer processes. Effects of mass and heat transfer, specific operating conditions need to be analysed and understood in depth. Such effects are specific for the reactor, processes and conditions used. Therefore, computational tools are ideal supporting analyses and comparisons of the kinetic data resulting from experimental studies in different set-ups and under ranges of conditions. At present the implementation of molecular models into simulations of technical processes remains anything but straight-forward.

Despite the general knowledge about these and other chemical shortcomings there was little development both in academic research and in funding projects that would acknowledge the substantial effort (and resources) that are needed to overcome these bottlenecks. Any initiative for digital catalysis science should take this into account as in the end it is the proven realization of novel systems with high performance that will convince the stakeholders about the usefulness of the new science paradigm.

4. Objectives

GeCatS aims to initiate and shape the transformation of catalysis research in Germany towards a sustainable digital future. Defining formats for experimental and calculated data, creating the required information infrastructure and involving all relevant users within the catalysis community will be the fundamental cornerstones of this concept. In a first step, the catalysis community is challenged to define common data formats and standardized testing and reporting procedures. Centralized initiatives will be created to set up, validate and operate the related information architectures in a sustainable way in the long term. The created infrastructure will enable all interested parties to use the benefits of digitalization-based research. This research will ultimately build a bridge between computational and experimental research in catalysis.

The implementation of a digital catalysis community in Germany will address all layers of the model outlined in Figure 4 holistically (onion model of digital hemispheres). Driven by the needs of the catalysis research institutions, both from academia and industry, and the actual users - the individual researchers - possible ways to improve their work will be identified, including type and amount of data, typical work flow, bottlenecks in data generation, handling, analysis, interpretation, conversion and storage. Based on this information, user friendly interfaces with a high degree of functionality will be defined and implemented. Such interfaces are intended to provide also access to tools for data analysis. Data will be stored in unified formats and derived from different sources, e.g. academic and industrial researchers, national organizations or publishers. Within the information infrastructure a core element are universal data bases that will serve as the cores for data storage. Options will be discussed to fully centralize such a concept, but also to allow individual institutions to build their individual pool of data and data tools in a way that remains interoperable with the overall concept and infrastructure.

On a technical level, the design of a shared information environment will be one essential core element. For such an environment, methods and approaches of organizing and labeling data, the extent of metadata, indicators for the origin and intended / permitted use of the data need to be agreed on. To reach many users, a related community needs to be created that facilitates the creation of software and algorithms, provides user support and works on aspects of data generation and handling according to FAIR principles. Defining best-practice procedures using different case studies will be key to improve the value and interaction with the catalysis research community.

The catalysis community is not alone in its quest to harvest the opportunities offered by a digital transformation. NOMAD represents the most advanced European initiative that follows the FAIR principles while addressing the challenges of unified data storage, retrieval and evaluation in the field of computational chemistry. The NFDI4Chem ini-

Executive Summary Core Statements “Objectives”

GeCatS aims to initiate and shape the transformation of catalysis research in Germany towards a sustainable digital future for the benefit of the entire community of researchers active in catalysis in academia and industry.

Within the information infrastructure the core element is a central data repository for catalysis related data. Parts of the infrastructure should be fully centralized, but also to allow individual institutions to build their individual pool of data in a way that the architecture remains interoperable and functional in conjunction with the overall concept.

The process of transformation should be driven by the needs of the catalysis community and the actual individual users, both from academia and industry, and allow all possible ways to enable and constantly improve the work-basis, including type and amount of data, typical work flow, bottlenecks in data generation, handling, analysis, interpretation, conversion and storage.

tiative (“Nationale Forschungsdateninfrastruktur für die Chemie”) outlines solutions for a sustainable access to research data in chemistry. Expert panels of the DFG (“The Digital Turn in the Sciences and Humanities”) discuss the influence of rapid digitalization in a broad range of different research fields. Synergistic collaborations with existing and emerging initiatives will be a key factor for success also in catalysis. The experience in information architecture development gained by TIB Hannover, MPG or Helmholtz Society could help in guiding the efforts of the catalysis community. Interdisciplinary links to informatics and mathematics sciences will be essential in this endeavor. The identification of new academic partners in informatics and mathematics, working on common catalysis projects and be attractive for young researchers e.g. chemo- or bioinformatics or software engineers should be a part of the strategy.

However, the listed examples also make it clear that measuring, storing and accessing data require the creation of a sustainable, reliable and lasting infrastructure including significant personal budget for experts. Such infrastructure will require also sustained funding schemes that can operate on a time scale much longer than typical single-user research projects.

Moreover, the generation of high-quality experimental data requires significant time, experience and dedication, which should be reflected in attractive funding opportunities also for the experimentalists.

A wide diversity in interests of the individual members of the catalysis community needs to be acknowledged. While the described digital transformation offers a multitude of promising opportunities, also the challenges faced by the individual researchers, research institutions, funding organizations, industry and the whole catalysis community need to be addressed to allow widespread adoption, acceptance and contribution. These challenges go way beyond the technical issues outlined above. All the listed actors need to be able to develop a sustainable mode of operation or business model, where ownership and access to data and algorithms need to be defined, and both academic and commercial interests can be satisfied in a balanced way. Apart from the expectation that digital catalysis will be beneficial for catalysis as a science, it is also expected that a digital catalysis scene will open new business opportunities beyond established economies. Funding schemes need to address and foster new business initiatives adequately in connection to digital catalysis.

While possible gains for the society are clear and obvious, the personal perspective of individual researchers and the community perception within such a system will be essential to success. If researchers are expected to act according to the FAIR principles, sharing their data, only adequate rewards and advantages will ensure a motivation that produces high quality data. New rewarding schemes or infrastructure initiatives focused on “high quality data”, publication credits by other researchers for the use of data, rewards for inventions made based on such data and other aspects need to be addressed. These aspects, if addressed adequately, will lead to an embracing by the community and make the approach and the cultural impetus required “self-igniting and “self-propellant”.

5. Next Steps

GeCatS is convinced that a national digitalized community of scientists active in the field of catalysis will be able to tackle national and global societal challenges in a much more efficient manner compared to the capabilities of the community in catalysis in its present state. In order to allow for transformation of the scientific community in catalysis GeCatS is convinced that an implementation plan is required which ensures that the technological basis is laid as solid foundation in a sustainable manner. Any plan must respect the fact that a fundament for digital catalysis, involving an information infrastructure and different pillars of data formats to be used by the community must be conceptually designed in a fashion that they are functional and widely accepted not only for a limited period of time. The whole information architecture as well as the data formats and must be evolved respecting community consensus, so that the entire approach and the digital community will have a future the community itself can shape.

As described in the chapters before, the core technical elements, information architecture and data formats, play a vital role and need to be introduced based on a participation model which allows community consensus. Additionally, it is vital that the scene of scientists is encouraged to use the technical backbone, to explore its possibilities and to develop it further. Experiences in digital transformation processes in industry have shown that such transformation processes require especially resources devoted to the development and education of human capital, apart from resources devoted to technical infrastructure, information infrastructure design and development efforts in data formats and digital tools for data analysis. An important part of the proposed implementation plan of GeCatS is therefore to not only foster the implementation via national funding for information infrastructure and development of data formats, but to go beyond and to invest in human capital to build a strong and educated national community by shaping the future by new educational initiatives at universities.

To achieve this, GeCatS supports to establish a long-term project covering the phases of definition creation and initial use of digital catalysis science. Such a project will need to be structured in sub-projects covering technical organizational and educational aspects. It will further need a branch of experimental data library generation. Finally, the project needs developing a roadmap towards institutionalization of these elements that require continuous existence. The whole process should be user-driven and follow the principles of stakeholder participation, from both academia and industry, in all its elements. In conjunction with this initiative it is also vital to work on

Executive Summary Core Statements “Next Steps”

GeCatS is convinced that a national digitalized community of scientists in the field of catalysis will be able to tackle national and global societal challenges in a much more efficient manner compared to the capabilities of the community in catalysis in its present state.

For the transformation of the community of researchers in catalysis GeCatS proposes a stepwise approach based on four program phases hosting lighthouse projects via one common publicly funded national program. Resourcing needs of the national program is required to cover the resources of all four phases. The program design needs to specifically include and allow for collaboration of all parties of relevance from academia and industry.

A sustainable approach must be taken that ensures that a national digital catalysis platform is set up in a way that it can be entertained and scaled in cooperation and dialogue with the community even beyond a dedicated phase for setup and use of the system and the infrastructure within the framework of a national program.

The cultural change in the scientific catalysis community must be accompanied, fostered and strengthened through the development of educational formats for future generations of scientists, which need to be educated and trained for and within a digital catalysis community.

models for an implementation plan in the national educational system, so that future generations of scientists working in the field of catalysis do have a smooth start regarding the skillsets acquired during their education. From a current point of view, it is also important to emphasize that the time scale until a full implementation and the final goal of a fully digitalized national scene in catalysis can be reached is expected to be on the order of a decade, possibly longer. As a target for such a time frame it is anticipated that the information architecture will have become an indispensable tool of a digitalized community in catalysis on a national and international basis.

5.1 Stepwise Approach

For the transformation of the community of researchers in catalysis GeCatS proposes a stepwise approach based on four phases (see figure 6). The stepwise approach has the advantage that progress can be monitored, and measures can be taken to ensure that overall project success is achieved. In the following the four phases will be illustrated and discussed:

Phase I: Foundation Phase

In the initial or foundation phase the technical fundament for a digitalized national catalysis scene will be laid. The core questions that will be addressed will be circled around the two main components that are required to arrive at an operational system. Core questions around the adequate information architecture must be addressed. Especially here an interdisciplinary dialogue in between groups active in experimental catalysis, theory & and simulation with mathematics, informatics and specifically groups being active in the field of information architectures is vital to lay a solid foundation for future work and the upcoming phases of the project. The main role of members of the catalysis community in this part of the endeavor is to give input on the functionality of the system to be achieved later. As parts of the system will require cloud-based solutions and others supercomputing facilities it is vital that in this part of the project a powerful hosting organization is identified. A variety of organizations could be potential hosting organizations for an infrastructure of such a complex format; potential examples of hosting organizations could be institutes within the Leibniz Society (like the TIB), the Helmholtz Society, the Max Planck Society, Societies like DECHEMA or a combination of organizations represented by a "Gemeinnütziger Verein". As addressed in the previous chapters an emphasis must be made on the questions of adequate data formats regarding catalysis based on the five pillars illustrated. Apart from the consensus of the community for ontology and taxonomy adequate formats must be identified that also ensure maximum functionality of any system run within the information architecture. In addition to the pure technical developments and discussion on the consensus of data formats and content, as a follow-up exercise the functionality of the system must be validated. During that validation phase several core functions must be validated by a defined group of users in close contact and collaboration to system developers

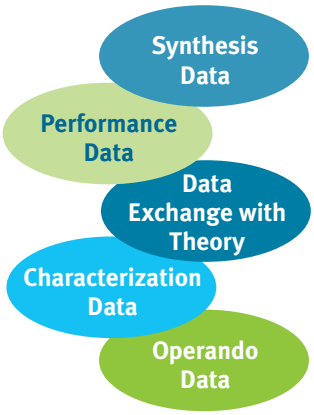
Phase I "Foundation"	Phase II "Excecution"	Phase III "Implementation"	Phase IV "Operation Mode"
	Series of parallel "Lighthouse Projects" Ambitious development challenges in catalysis with potential high socio-economic impact Ambitious development targets with regard to application of new data science methods in catalysis	Series of parallel Endeavours targeting at the implementation of the new approaches in educational formats which lead to new generations of well trained scientists	Post Project Phase: Information architecture is entertained and expanded and used by community
Data formats are created and information architecture is established	**Lighthouse projects are fostered by digital toolsets developed**	**Digitalization finds its way into education and daily research**	**Digital platform is standard tool in research and education**

Figure 6: Project execution plan for digitalization roadmap in catalysis.

5. next steps

and architects. The ambition is to address a broad field of expertise within the field of catalysis in a cross-disciplinary effort that allows to illustrate functionality and generate a first generation of users which are familiar with the system. The first checkpoint for *Phase I* will be "Data formats are created on community consensus, information architecture is established, functional and validated by first user group". Experimental benchmark studies to knowledgeably compare and combine work done with different equipment and conditions is explicitly included in Phase I. This experimental validation is an analogous approach that has proven successful to that of "test sets in quantum chemistry".

Resources required within this phase will be necessary for the following topic clusters:

- resources for setup, structuring and entertainment of information architectures
- resources for development of accepted and adequate data formats
- resources for creation of data entry points and entry pathways and procedures (data-conversion agents)
- resources for validation of system architecture and functionality

Phase II: Execution Phase

The main target of the *Execution Phase* is to get a maximum number of users involved inside and outside of lighthouse projects to actively work with the digital platform. A cluster of „Lighthouse Projects" homebased in one common publicly funded national program should have two common distinctive features: they should all address an ambitious development challenge in catalysis with potential high socio-economic impact and secondly at the same time have ambitious development targets with regard to application of new data science methods in catalysis using the established platform.

The selection and the monitoring process of the lighthouse projects is an important element, which needs to be implemented in the common publicly funded national program. It is evident that even a substantially resource equipped national program cannot fund all initiatives for lighthouse projects desired by the community – therefore the information architecture it is recommended to make the information architecture available to the public after *Phase I*, so that a broad acceptance within the community is fostered.

The work-attitude within the "Lighthouse Projects" with the digital platform should be focused on three major aspects:

- harnessing the digital platform to harvest its benefits and to establish a new work culture in the field of catalysis on a community basis.
- development of new methods based on sharing and open access policies under the FAIR principles

and finally

- constant improvements of system functionality and architecture based on user and community experience.

It is expected that the lighthouse projects will not only lead to a step change in work culture but also to dramatic changes in development pace and the level of insights gained within the topics addressed. One of expected values generated apart from the cultural step change are the competitive advantages in the respective fields in catalysis addressed, a fact that should also materialize in technological developments that find entry into commercialization.

The checkpoint for Phase II will be "national catalysis community brings information architecture in lighthouse projects to bear value on topics of high socio-economic relevance".

Resources required within this phase will be necessary for the following topic clusters:

- resources for execution of lighthouse projects
- resources for transformational exercise of catalysis community in using the digital platforms on a project basis
- resources for development of new digital tools and methodologies available to the community

Phase III: Implementation Phase

Phase III is focused on fortifying the new approach by finding entry into step-change in research formats and even more importantly into education. The ambition here is to equip new generations of researchers with the necessary educational background to allow them a quick start with the information architecture in place.

Phase III should be organized in series of parallel endeavors targeting at the implementation of the new approaches and appropriate content in educational formats which lead to new generations of well-trained scientists. It is desirable that within the program a certain governance is executed regarding details of entry into educational formats to ensure nationally harmonized approaches.

The checkpoint for *Phase III* will be "digitalization finds its way into education and daily research on broad basis through scene".

Resources required within this phase will be necessary for the following topic clusters:

» resources devoted to development of educational formats for a future generation of scientists trained for a digital catalysis community

» resources for the development and steering of national educational programs and initiatives

Phase IV: Operation Mode

The last phase of the envisaged program focusses on transferring the information architecture into a constant digital operation mode that ensures functionality for a growing community of researchers on a national and international basis in harvesting the benefits in up-to-date research formats and education.

The core mission in this phase is to entertain and expand the information architecture by a host organization and used by the digital catalysis community. As addressed in Phase I, the question of hosting has already to be solved at an early point in the program to ensure that in the final phase the future of the information architecture is not compromised by gaps in system entertainment of failures in scalability. Therefore, an organization must be identified at an early point being capable of ensuring a future of the platform.

The checkpoint for *Phase IV* will be "national digital catalysis platform is entertained and scaled in cooperation and dialogue with the community".

Resources required within this phase will be necessary for the following topic clusters:

» resources devoted to system entertainment

» resources devoted to dialogue with the community

6. Conclusion

The science of catalysis is about to experience a disruptive change in its core paradigms. Several experimental and theoretical developments allow for the first time to describe catalysis in its whole functional complexity without having to make simplifications that break the relation between model and performance systems also applied in an industrial environment. This is valid in all types of catalysis: heterogeneous, homogeneous, bio-, electro- or photo-catalysis. The central development that brings together all these developments is digital catalysis science allowing a new perspective on catalysis. This paradigm has the potential to get closer to the dream of designing catalytic systems "from scratch". The system approach linking the desired reaction with its reaction network, the catalyst, as dynamic entity, and the operation conditions is essential to obtain the desired predictions. These include a physics-based description of all relevant elementary steps such as changing chemical structure and transport of matter and energy. In Figure 7 the relation between contributing fields enabling digital catalysts science its consequence for designing catalytic systems and the massive need of several technical-industrial challenges with societal dimensions for this breakthrough innovation in catalysis.

GeCatS as the platform in Germany for catalysis science established by learned Societies pushes for a forward-looking combined research action to develop, implement and operate this new paradigm. The combination includes stakeholders from academia, chemical industry, information and publishing industry and funding institutions. Within academia the disciplines of physical chemistry, organic chemistry, inorganic chemistry, bio-related sciences and chemical engineering as well as physics, mathematics and informatics are asked to collaborate and leave behind the divide between catalysis disciplines.

The community has started the process with stakeholder meetings, and a GeCatS Infoday was held in November 2018 the results of which form central elements of the present document. If Germany with its long tradition in creating breakthrough innovations in chemical industry wishes to continue this role then a powerful and lasting effort is necessary to move forward with the well-recognized transformation of the research workflow in catalysis science. GeCatS seeks the support of politics and of funding organizations and wishes to send out a strong signal that the opportunity to enter this disruptive development is here and now.

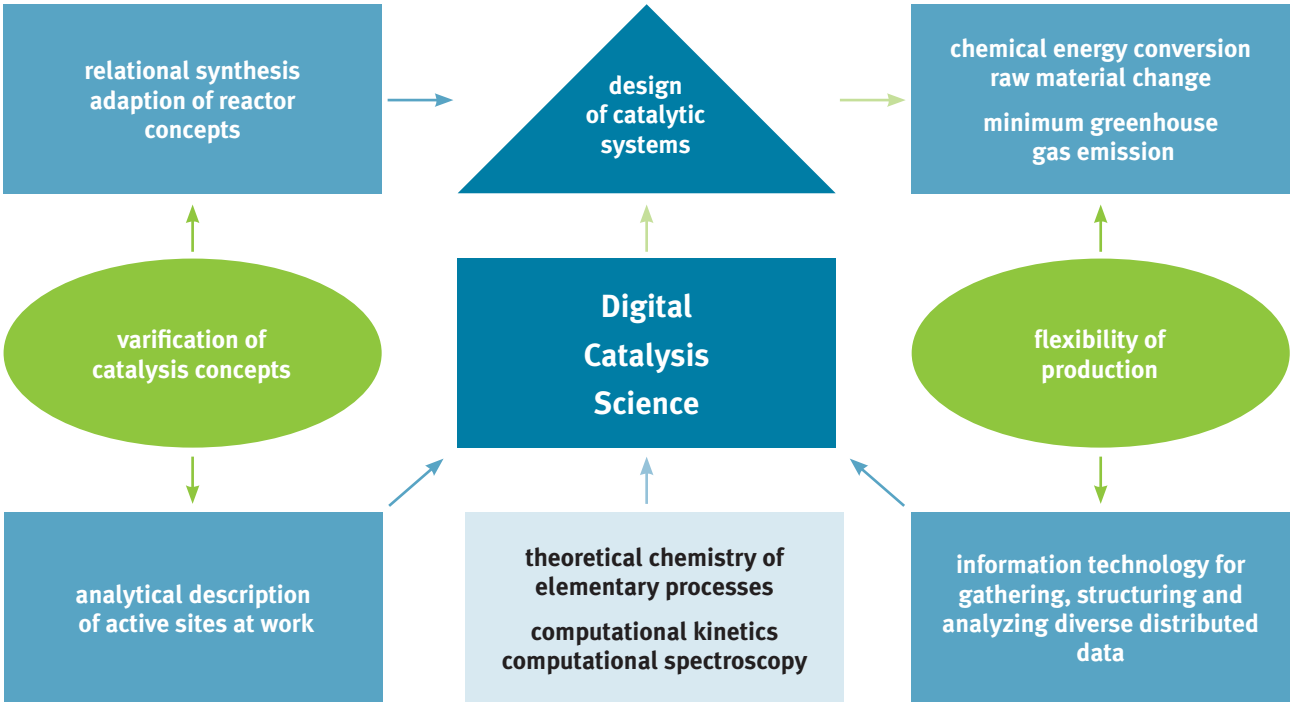


Figure 7: Enabling fields for digital catalysts science and its consequences for science and technology

LIST OF CONTRIBUTORS

Dana Demtröder, DECHEMA e.V.	Member Task Force, Author, Editorial Team
Olaf Deutschmann, Karlsruher Institut für Technologie	Author
Bernd Eck, BASF SE	Author
Robert Franke, Evonik AG	Member Task Force
Roger Gläser, Universität Leipzig	Member Task Force, Author
Lukas Goosen, Ruhr Universität Bochum	Author
Jan-Dierk Grunwaldt, Karlsruher Institut für Technologie	Member Task Force, Author
Ralph Krähnert, Technische Universität Berlin/BASCAT	Author
Udo Kragl, Universität Rostock	Member Task Force, Author
Walter Leitner, Max-Planck-Institute for Chemical Energy Conversion	Member Task Force, Author
Gerhard Mestl, Clariant AG	Member Task Force, Author
Karsten Reuter, Technische Universität München	Author
Frank Rosowski, BASF SE/BASCAT	Member Task Force, Author
Ansgar Schäfer, BASF SE	Author
Matthias Scheffler, Fritz-Haber-Institut Berlin	Author
Robert Schlögl, Fritz-Haber-Institut Berlin, Max-Planck-Institut für Chemische Energiekonversion	Author
Ferdi Schüth, Max-Planck-Institut für Kohlenforschung	Author
Stephan Andreas Schunk, hte GmbH	Member Task Force, Author, Editorial Team
Felix Studt, Karlsruher Institut für Technologie	Author
Kurt Wagemann, DECHEMA e.V.	Author
Peter Wasserscheid, Friedrich-Alexander-Universität Erlangen-Nürnberg	Member Task Force
Christoph Wöll, Karlsruher Institut für Technologie	Author
Dorit Wolf, Evonik AG	Member Task Force

PUBLISHER

German Catalysis Society (GeCatS)

Responsible for content under the terms of press legislations:
DECHEMA e.V.
Dr. Andreas Förster
Theodor-Heuss-Allee 25
60486 Frankfurt am Main
Germany

Publication date: March 2019

DECHEMA
Gesellschaft für Chemische Technik
und Biotechnologie e.V.
Theodor-Heuss Allee 25
60486 Frankfurt am Main

Telefon: 069 7564-0
Telefax: 069 7564-117
E-Mail: info@dechema.de