Journal of Business Chemistry

Vol. 15, Iss. 2

www.businesschemistry.org

06.2018

The academic journal for management issues in the chemical industry

Published on Behalf of the Institute of Business Administration at the Department of Chemistry and Pharmacy, University of Münster and the Center for Industry and Sustainability, Provadis School of International Management and Technology

Aim and Scope

The Journal of Business Chemistry (JoBC) focusses on current developments and insights at the intersection of management and chemistry, biotechnology or pharmacy.

The JoBC provides an international forum for researchers and practitioners in companies, research institutes, public authorities, consultancies or NGOs to present and discuss current challenges as well as potential solutions in an interdisciplinary manner. Thus, the JoBC aims to foster the dialog between science and business, to support management practice in the chemical and pharmaceutical industry and to indicate where further research from academia is needed. The JoBC offers high quality publications with academic standards, a fast publishing process and global reach. With this multidisciplinary and boundary-spanning approach, the Journal of Business Chemistry intends to become the leading journal for decision makers in the chemical and pharmaceutical industry.

Publisher

The Journal of Business Chemistry (ISSN 1613-9623) is jointly published by the Institute of Business Administration at the Department of Chemistry and Pharmacy (University of Münster) and the Center for Industry and Sustainability (Provadis School of International Management and Technology). It is published every four months as an open access journal.

Subscription

Online-Subscription is possible at:
subscription@businesschemistry.org.
The articles are available for download free of charge at www.businesschemistry.org.

Abstracting and Indexing

The Journal of Business Chemistry is covered by the following abstracting and indexing services:

- EBSCO Publishing
- Hamburg Institute of International Economics
- German National Library of Economics
- Chemical Business NewsBase (CBNB) by Elsevier Engineering Information, Inc.

Copyright

Contents

ISSN 1613-9623

Letter from the Editors

Digitalization in the chemicals: How to succeed the digital transformation

Hardly any other topic dominates the current disourse in the chemical industry like digitalization. The digital transformation is in full swing. Production facilities are becoming more intelligent, processes are shifting from analog to digital, algorithms creating valuable insights and new, digitally enabled business models are evolving. However, the goal is not a digital transformation per se but rather a business transformation: harnessing digital capabilities to transform traditional products and services into reimagined value propositions in the digital era. In oder to achieve future success, all companies in the chemical industry and related sectors need to embark on a digital transformation journey. The present issue of the Journal of Business Chemistry sheds light on the topics mentioned with a commentary and three contributions in the practitioner's section.

What's the role of chemical distributors in the course of digitalization? In his commentary "Digitalization, chemical distribution and the chemical value chain" Till Knorr shares his thoughts on what chemical distribution could look like in the future. Although digitalization could make some tasks from chemical distributors obsolete, there is no digital substitute for the physical delivery in sight. Marketplaces will prove to be central orchestrators and consequently play a dominant role in value creation processes.

In the digital economy marketplaces and online shops are key channels to market products and services. Chemical companies already set up their own digital platforms. However, commercial buyers of chemicals were still missing a website or a tool in order to compare the offer from different suppliers at one central place. In the article "The digitalization of marketing and sales in the chemical B2B sector" Rune Koehn presents a case study of a startup that developed the first metasearch engine for chemicals.

Companies face intense pressure to keep themselves at the leading edge of how to utilize data for long-term financial and digital success by applying advanced and complex technologies such as artificial intelligence. Effective technological solutions require a well-structured approach. Robert Jenke describes in his article "Successful data applications: a cross-industry approach for conceptual planning" the CRISP-DM model as an established framework to guide projects regarding data usage in the process industry.

The biologization of traditional industries such as chemicals, pharmaceuticals and life sciences has become an ongoing trend. The article "Quantitative biology - a perspective for the life sciences' way into the future" by Holger Wallmeier gives a broad overview about quantitative biology and its implications for scientific and industrial research. Quantitative biology is best described as the combination of biology, statistics and mathematics, and it relies on special expertise. Data scientists will have a key function in future research but require further understanding in the fields of the origin of the data.

Please enjoy reading the second issue of the fifteenth volume of the Journal of Business Chemistry. We are grateful for the support of all authors and reviewers for this new issue. If you have any comments or suggestions, please do not hesitate to contact us at contact@businesschemistry.org.

Thomas Kopel (Executive Editor)

Bernd Winters (Executive Editor)

Commentary
Digitalization, chemical distribution and the chemical value chain

Till Knorr*

* Till Knorr, EXPACON GmbH, Am Wildwechsel 6 A, 82031 Grünwald, Germany, tknorr@expacon.com

DOI: 10.17879/38129710787; URN: urn:nbn:de:hbz:6-38129712134

Do you still remember ChemConnect, CheMatch, cc-chemplorer and the like? More than 15 years after these companies set out to disrupt the chemical value chain, the topic is back. While most of the then startups have failed (Elemica being the notable exception), today's efforts might fare differently. In this article we focus on distributors who link producers and (usually) small customers. However, some of our thoughts affect the complete value chain.

We start setting the scene for digitalization, then step back a little to discuss some common misconceptions on the role of a distributor, followed by ideas on properly managing the relationship principal/producer vs. distributor. Finally, we come back to digitalization by discussing electronic marketplaces and their impact on the chemical value chain.

1 Digitalization in chemicals – getting rid of distributors for tail-end customers and unknown territories?

When talking about digitalization, disruption is all the rage. Disruption of business models, disruption of value chains, disruption of everything that has to do with customers. So, with all major chemical producers talking about digitalization projects in the range of two- or three-digit millions, disruption of the chemical value chain is already underway. But what will be affected and – focus of our thoughts – is that the end of chemical distribution?

First let's start with the good news for all the companies that produce, distribute or process chemicals. We will continue to live in houses, eat whatever we like (or what an influencer promotes), dress up and use interfaces with the digital world. In our modern world all of that requires chemicals. Hence, whatever digitalization is going to change – chemicals as building blocks or as ingredients that improve performance, safety and durability of things of our physical world are going to last. And thereby its producers as well as processing companies. But what about distributors?

Digitalization will definitely change all processes (and businesses) which deal with information. Whatever can be digitized, surely will. For the market side of the chemical value chain this translates into direct communication between chemicals processing enterprises (customers) and producers. Whenever a customer's tank or warehouse space runs empty, in a completely digitized world this (already existing) piece of information could directly feed the production (and marketing) plan of the producer, cutting out all the intermediaries, i.e. the chemical distributor. This idea probably resonates well with many a business head who regards distributors as a necessary evil with no other purpose than pocketing the rightful margin of the producer.

However, if we take a closer look at the value chain AND at the wishes of the customers, there is more to a distributor than just passing through information. Firstly, the distributor not only acts as demand bundler in the above sense towards the producer, rather it bundles the demand for various chemicals/supplies towards the customer. Accordingly, a customer is not interested in getting chemicals from several producers (meaning several deliveries at its warehouse), if he can get them bundled in one delivery from one distributor. Secondly, the distributor takes care of the physical delivery of the chemicals, if required including repackaging in unit sizes smaller than the producer offers. Thirdly, with some specialties, the distributor acts as technical "first level support", freeing time of technical sales staff of the principal. And fourthly, a distributor can and should act as a trend scout and service provider (e.g. customs clearance) in markets too small for a producer/principal to serve. This latter function makes sense, if the distributor is already present in a certain market the principal wants to enter.

In a first and quick assessment, the market scout function and the technical "first level support" could indeed fall victim to digitalization:

- Market scout to artificial intelligence which-systematically collects all data on a certain market and then comes up with predictions on which sectors will grow at what speed and with which demand of certain machinery, chemicals, financing etc. However, this kind of information will then be available to ALL interested buyers and no longer confined to a one-to-one distributor/principal relationship.
- "First level support" to artificial intelligence and/or augmented reality. In case of problems with the use/application of a chemical, there might be solutions coming up with no human staff involved. E.g. an artificial intelligence interface with search options for solutions of the problem within a huge database of reported problems. Or, if machinery or the like is involved, augmented reality might guide through the problem-solving process.

For the two other roles of a distributor in the value chain there is no digital replacement in sight – the bundling of products for the customer and its physical delivery. That's the good news. The bad news is, there are different players operating in these areas which might qualify as the future winners, hence have to be watched: chemical distributors, (chemical) logistics companies, amazon/Alibaba or similarly powerful internet giants.

We do not reckon incumbent chemical companies as serious competitors in this arena, due to the customer need of bundled deliveries (often more than one supplier) and physical delivery. While the information flow from customers to producers/principals will work seamlessly, the complexity of managing all the product needs in addition to one's own products is going to kill any business case. Especially since the logistics requirements usually differ in size and volume, asking for infrastructure investment at the wrong (tail) end of the customer base.

If a business head of a producer still thinks the chemical distributor a redundant money-grubber, a different understanding of the business model might help. But that's topic of our next section.

2 Chemical producers and chemical distributors – often a story of many misunderstandings

Over the last decades we've seen distinct developments on both the producers' as well distributors' side. Whereas distributors are increasing geographic spread (consolidation) and application technology know-how, producers are focusing on core businesses while divesting others. These trends are leading to distributors covering ever larger geographical areas and producers concentrating on a limited range of products/applications with (in total) smaller sales teams. Those two developments should generate various opportunities for mutually beneficial collaboration. However, it is surprising to see how often little knowledge about the other side exists and how many prejudices still prevail. One minor reason might be a hazy positioning as commodities/specialties/full line distributor. Already difficult for producers, a distinct focus on commodities, specialties or both is extremely hard to achieve for distributors. (Due to the almost endless variety of chemicals and their applications sold to almost all industries.)

Nevertheless, distributors can work towards a better recognition at producers. Usually, producers see distributors as THE channel to serve small- to midsized customers. Whereas this view is clearly understandable, it is a little surprise that often the physical realisation of the distribution is not recognized. In contrast, distributors see their role without exception in exactly this physical distribution and, to a large degree, in the bundling of products for customers. Explanations for the different views probably come down to different perspectives due to one's own day-to-day work. Whereas bundling represents the customer perspective of a distributor, which is (rightly) neglected by most producers, the wish for serving and developing small customers of a producer is only part of the self-image of distributors. These different conceptions offer an opportunity to start discussing what a producer expects from a distributor and what a distributor can offer.

Merely seeing distributors as an extended sales arm means missing chances of benefitting from their physical infrastructure (e.g. labs, warehouses, etc.) or their knowledge about untapped markets. The same applies to know-how. Producers (principals) require distributors to be just as knowledgeable about applications as they are about products; distributors, however, value their application know-how much higher. And rightly so. Solving customers' problems offers more business opportunities (and stabilizes relations) than pure product

Figure 1 Typical activities of a specialties distributor (source: own representation).

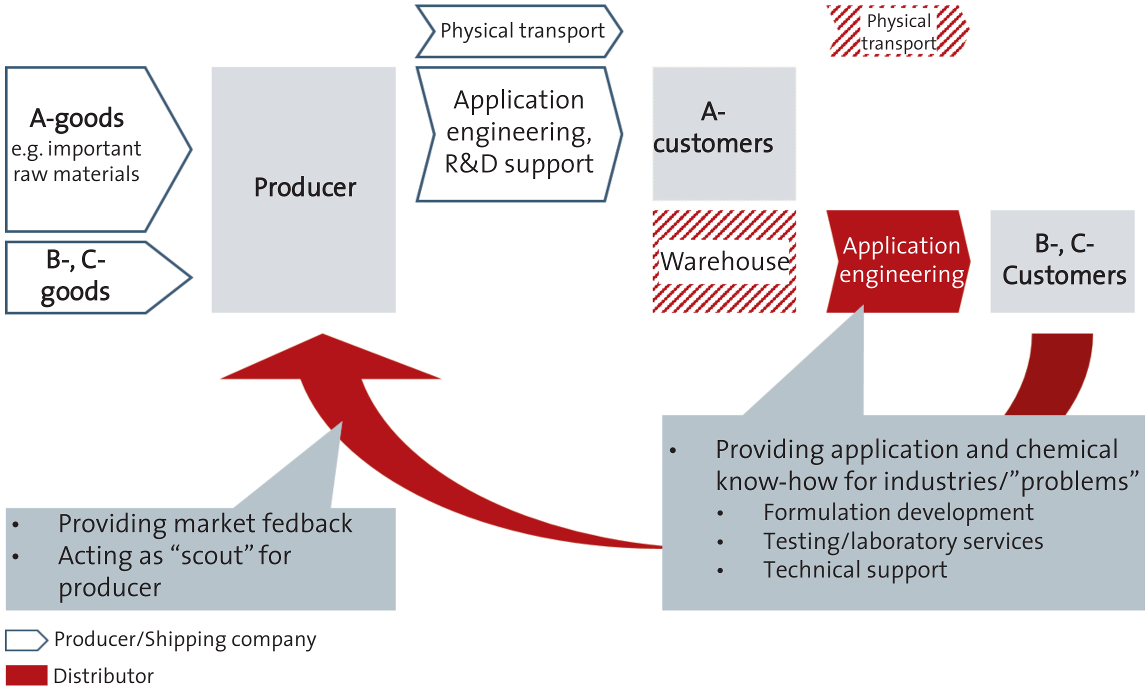
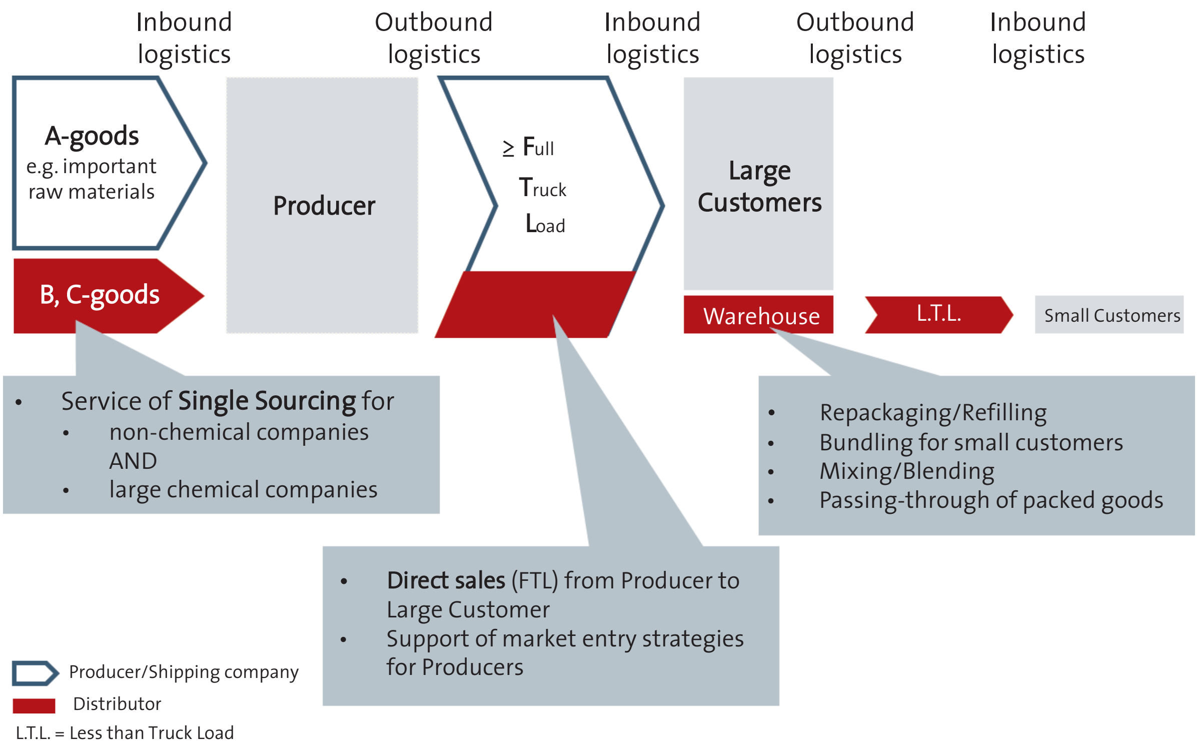


Figure 2 Typical activities of a commodity distributor (source: own representation).



know-how. Therefore, principals should expect their distributor to be an application expert and jointly discuss applications and their development. But they should also reward them accordingly.

Equally, distributors should, to a larger degree, accept this role as an extended sales arm. Even if own product lines could prove highly successful, the exceedingly larger part of the business will result from the sales of principals' products. Therefore, winning over the right principals with appropriate offerings must be at the centre of attention. The design of respective offers should also consider those wishes of principals on which distributors seem to focus less. This is presence in countries which are no longer or not yet served by principals, offering lab capacities for dedicated applications as well as having low cost structures. However, it is not quite clear as to whether principals' wishes (such as reducing administrative cost) have been achieved already and simply not been talked about, or whether this is another opportunity for distributors to enhance their offerings by services that aim at exactly this goal.

3 Systematic approach to distributor management

As long as producers/principals on the one and distributors on the other hand have a blurred picture of each other, no mutually beneficial relationship will evolve. Therefore producers/principals should be able to clearly address what they expect from a distributor. Not an easy task, if the overwhelming majority of one's business (and attention) dictates different priorities. However, for a systematically run business, successful distributor management can be done along a few simple rules.

It all starts with the "ordinary" business strategy, including focus products, target customers, regions etc., followed by a respective resource allocation to make the plan work. *(Step 1: Build on your business strategy)* Business as usual. And as long as key customers or key regions are concerned – planning should be easy. But, whenever one of the following situations turns up, a distributor might be of help *(Step 2: Identify relevant areas)*:

- a new market shall be developed
- a large number of small customers ("tail end")
- a market has become too small to justify full sales resources
- customers ask for complexity-increasing additional services

And, the classic rule of thumb, always think about handing over the last 20% of customers, since they usually don't contribute to profits.

Next, search for distributors with expertise in the required applications or regions. Big full-line distributors are always a choice, but there are several specialists in a variety of fields. When discussing with the selected distributors, a combination of the above-mentioned situations might be addressed. *(Step 3: Select the right support)*. However, besides distributors there are different options. These include using agents, sharing staff with other businesses or dropping the business altogether.

Assuming, principal/producer and a distributor agree upon working together, the most difficult issue is to manage the business relation in a trustful manner. *(Step 4: Manage the relationship)* It is important to clearly define targets and the way to measure their fulfilment as well as to agree upon the depth of information exchange. Even more important is the definition of the most critical issues in such a relationship: how to proceed if a market or a customer grows above a critical threshold or giving technical and specialist support/training. This is the last moment when both parties can clearly announce and write down their particular expectations from the relationship. It is understandable if a producer/principal wants to take back customers grown above a certain threshold – but it is also understandable that the distributor who has put all his effort in growing this customer wants a reward for this achievement. Possible solutions could be a joint discussion with the customer about the best way to serve him in the future and/or some profit sharing for a defined number of years to come between distributor and producer/principal. Thus, for a distributor it is quite helpful to recall from time to time one's positioning in the value chain. But this is no one-way-road. A producer/principal has to accept, that a distributor is managing complexity and needs a certain budget (i.e. margin) to do so. This has to be fixed in the contracts as well.

As long as all this is done right at the outset, both parties are able to develop a trust based relationship.

4 Electronic market places as sales channel to eliminate distributors for small customers?

As already described above, distributors have an important role in the chemical value chain. They serve small customers, act as market scouts, offer services and commission as well as refill/repackage products. Market places on the other hand try to match supply and demand, i.e. are – at least at the moment – mainly dealing with information flows only.

These information flows are important, since

Figure 3 Basic value chain - information flows (source: own representation).

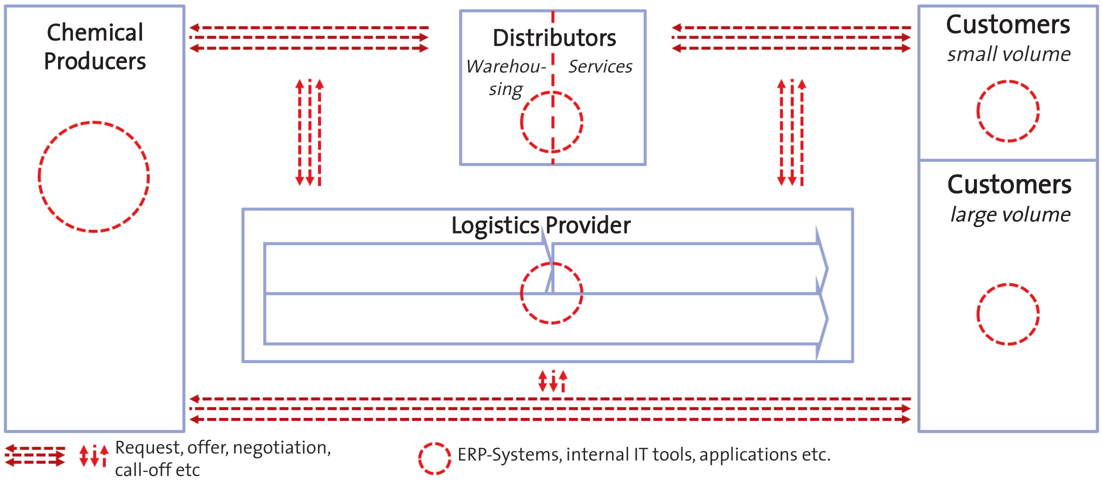
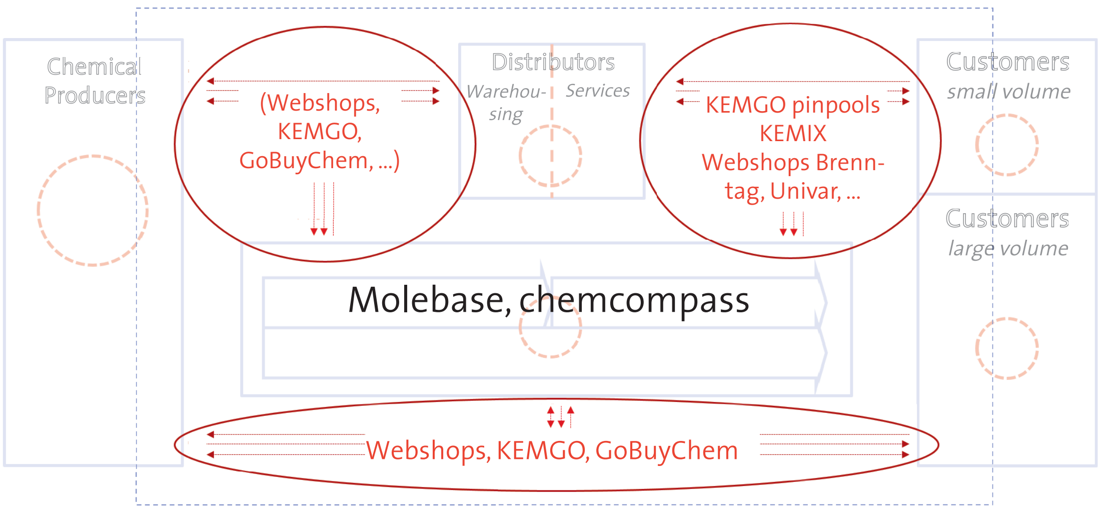


Figure 4 Basic value chain - evolving data processing options (source: own representation).



they are the basis for any sale to be made. A customer who wants to buy a certain chemical product needs to find suppliers. After finding them, either a deal can be made immediately (if volume/price quotes exist) or negotiations take place before agreeing on a deal. All that is pure information (or data) and can be dealt with by a market place. The moment the delivery comes into play, data is still required but no longer sufficient. Now physical product flows have to be handled – repackaging/refilling, commissioning and transportation. And that's the domain of distributors and logistics companies. Based on this understanding the "easiest" future scenario is market places managing the information flows, all other elements of the value chain remaining the same. However, due to the disruptive potential of today's technology, there are some more options imaginable.

1. Chemical producers serve small customers

It is easy to establish an information channel to small customers (webshop, market place), learn about their demand, offer a price and negotiate a contract. However, delivery has to be covered, too. Firstly, the respective infrastructure is required (tanks, warehouse space, filling stations, ...) and secondly, delivery costs for customers could increase dramatically if they have to buy from various producers instead of getting one pallet with several products from the distributor. Chemical producers shouldn't underestimate the complexity they would insource by tunnelling the distributor.

2. Market places offer additional distribution and logistics services

This is probably an option, market place operators have already in mind. However, it will prove difficult to find distributors willing to act as a mere contract refiller/warehouse provider for market places. They would have to leave parts of the margin with the market place and lose their customer contact. Not a promising option.

An additional difficulty – largely underestimated in today's discussions – is providing an electronic data interchange (EDI) interface to connect distributors, logistics companies, other service providers to the market place. Replacing/improving existing data exchange solutions will become one of the biggest obstacles on the way to a digitalized value/supply chain.

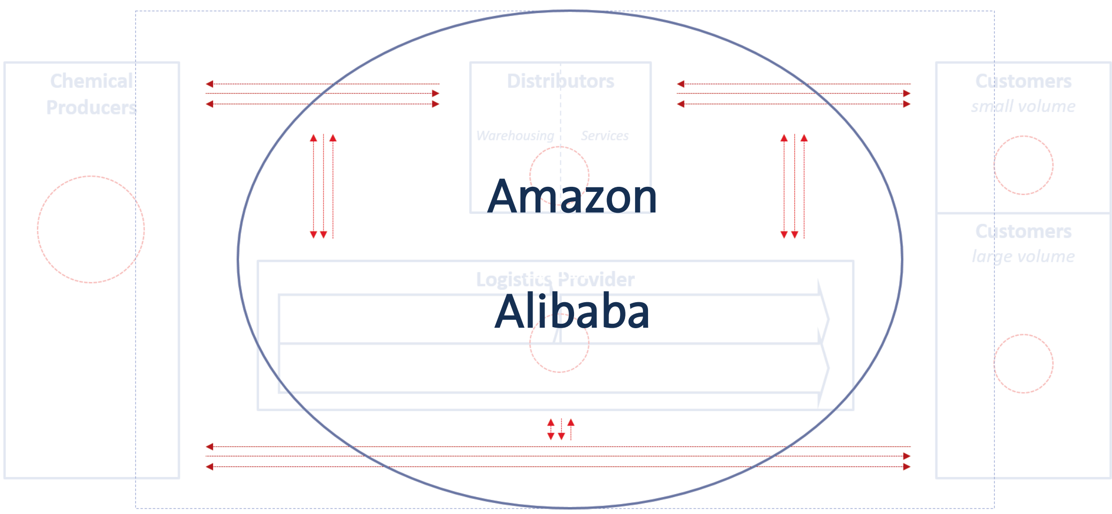
3. Amazon/Alibaba/... manage everything between customer and chemical producer

This is the most worrying option for EVERYONE in the value chain. First of all, market place operators will die since they are not able to compete with size and data processing power of the giants. Then distributors/logistics companies will both be downgraded to mere contract companies delivering warehousing/transportation services to the giants. After a while even chemical producers come under pressure. Due to their data analyzing power the giants are quick to spot best-selling/high-margin products and will, step-by-step, offer their private label products – downgrading even chemical producers to contract manufacturers. And finally, customers have no longer any bargaining/negotiation power since the data giants have total transparency on the supply/demand-situation and adapt prices accordingly.

(In addition, we could imagine distributors adding logistics capacities or logistics companies adding distribution capabilities. However, this is just a sub-case of the current "hardware set up" and not dealt with separately.)

To deal with today's uncertainty, we think it helpful to always distinguish between "hardware" (i.e. physical assets) and information or data. Digitalization starts with the latter and influences the "hardware". BUT, "hardware" will always remain a vital part of the chemical value chain. And in this sense distributors (i.e. their hardware) won't be eliminated. They might operate in a different setup, but their role in the value chain won't disappear. The much more pressing question for all players in the chemical value chain is, how to avoid the third option mentioned above.

Figure 5 Threat for ALL elements of the value chain - market places, distributors, logistics companies, customers and producers (source: own representation).



5 Surviving as part of the chemical value chain

In the last section we mentioned the scenario of a large internet giant occupying the space between customers and chemical companies – leaving producers, distributors, customers and platform operators worse off. Is there a way to prevent this?

Firstly, in case one of the giants is attracted by the chemical distribution business, that's probably it. With their huge advantages in gathering, processing and analysing data, they will outsmart any established player. Add the state-of-the-art logistics capabilities of an amazon (Alibaba?) and little more than the current customer base and deep knowledge on handling chemicals will remain. Given Jeff Bezos' infamous "Your margin is my opportunity", there would be one solution: become efficient and keep the margins razor-thin. Since this is definitely no solution for listed companies (and I guess neither for family-owned ones), is there any other way out?

One option is to create an amazon-inspired customer centric buying experience. This means creating a market place with a wealth of chemicals offerings – delivering transparency and comparability. In a second step, additional functionalities like safety data sheet handling or smooth payment processes could be included. However, there is one huge difficulty in such a marketplace becoming successful. It is the question of ownership.

- If a third party operates the market place, neither producers nor distributors will post a substantial number of offerings – due to the fear of losing valuable information AND a transaction fee.
- If one of the relevant producers or distributors operates the market place, the problem remains.
- Thinking it through, the obvious answer is – make it a jointly owned entity. By the way, this concept has already worked with Elemica. (In the early noughties almost all of the big chemical companies bought shares in the newly founded marketplaces).

Whether such an idea will get a chance, much depends on how likely a market entry by amazon, Alibaba etc. is seen. Opinions will vary. However, even if a joint platform comes into life, there are more obstacles to overcome. The biggest one is to ensure none of the owners interferes with the daily business of the platform. This has to be taken care of in the founding stages and requires codification in the charter. The same is true for a joint understanding of ways to generate revenues on the platform as well as profit distribution. In addition, the platform needs funding for the initial steps – a difficult issue among competitors. And as if that wasn't enough, in most cases the existing IT setup is not even near a status which allows for an integration into a platform. Neither are the internal processes "digital ready" nor is the required data readily available.

However, realizing the internet giants posing a threat to EVERYONE in the value chain should be motivation enough to start thinking about an industry wide platform. Whether the first steps are gone among distributors and customers or distributors and producers/principals finally doesn't matter. Doing nothing is likely to be the worst option.

References

Stratley (2012): CHEMICAL DISTRIBUTION - Findings at the distributor-principal interface.

Practitioner's Section

The digitalization of marketing and sales in the chemical B2B sector

Rune Koehn*

* chembid GmbH & Co. KG, Marie-Curie-Straße 1, 26129 Oldenburg, Germany
rune.koehn@chembid.com

DOI: 10.17879/38129710162; URN: nbn:de:hbz:6-38129710379

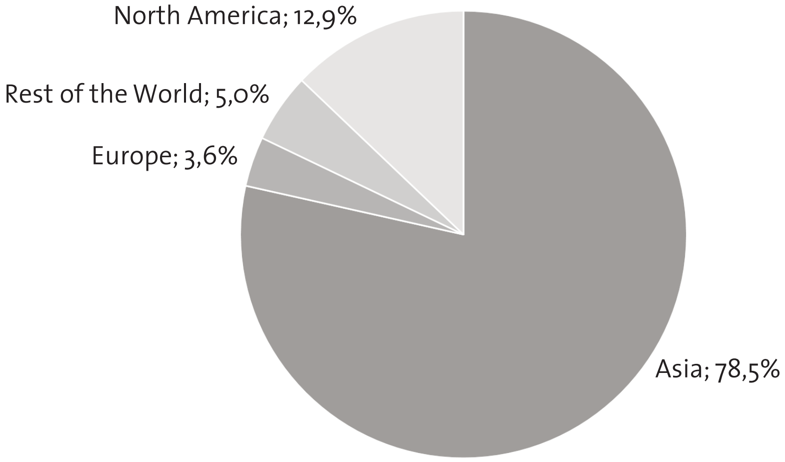
Marketing and sales in the B2B chemical sector are slowly making their way into the digital era. Especially in Europe the development is inhibited, as long established firms set the tone by listening to their best customer only, relying on deeply embedded routines and processes, and fostering their fear of cannibalization. In this article, the author outlines the current structures of sales processes in the chemical industry and presents a promising perspective for its digitalization by pointing out the great potential of disruptive newcomers in the form of innovative startups free from constraints of high-ranking incumbents.

1 Introduction

The ongoing digitalization does not stop at the Business-to-Business (B2B) sector. The internet is a key facilitator for entering the global market, but it also entails several challenges. While new technologies make it much easier to reach more customers or partners from all around the globe, it becomes more challenging to build stable partnerships (Samiee, 2008). The abundance of information available on the internet, makes global customers often better informed and more attentive to cost efficient offers of new players on the global market. Thus, suppliers are put under increasing price pressure and customer service and global networks become increasingly relevant (Matthyssens et al., 2008).

This article sheds light on current issues the European B2B chemical sector is facing. Digitalization and globalization are influencing the industry and presenting it with the challenge of successfully mastering the step towards the digi-tal era. The fields of research, production and processing have benefited from new, digital technologies from the beginning, while e-solutions for the sales sector seem to develop much slower (VCI and Deloitte, 2017). The aim of this article is to draw the attention of marketers, practitioners and managers of the chemical industry to the possibilities a holistic approach to B2B e-commerce offers for the chemical sector. It will begin with a description of the current situation of the B2B sales sector in the European chemical industry and inhibitors for incumbent firms to create disruptive innovations themselves. In chapter three, the article portrays the ongoing processes of digitalization, i.e. the development of e-commerce solutions for chemicals, claiming that the mere shift to digital sales solutions is just the first step to an efficient B2B sales concept. Afterwards, the disruptive potential of innovative startups and their impact on the industry will be outlined. As an example of a young spin-off on its way to innovate the chemical B2B sales sector, the author proceeds with a portrait of the German company *chembid*, which is the first firm to introduce an all-in-one platform solution for the global B2B market for chemicals and related services. The concept, functions and technologies of the platform in detail as well as the advantages of such holistic approach to B2B e-commerce for the chemical industry will be explained and an overview of the main benefits the 'chembid concept' offers for commercial sellers as well as buyers of chemicals is given. To conclude, we will summarize the key points of our argumentation for a holistic approach to B2B e-commerce in the chemical B2B sales sector.´

Figure 1 Shares of regional B2B e-commerce gross merchandise volume (GMV) in 2017 (source: Statista 2017).



2 Marketing and sales in the European chemical industry

Considering the global revenues from e-commerce, i.e. the use of the internet to conduct business transactions (Terzi, 2011), the trend is obvious. Between 2013 and 2017 the sales volume was expected to increase steadily. Last year, it was predicted to reach an estimated generalized mean value of US$ 7.661 million. Especially the regional distribution of this number, displayed in figure 1, reveals a clear picture: 78.5 % are turned over in Asia, mainly China, and 12.9 % in North America, while Europe lags behind with only 3.6 % (Statista, 2017). With the EU being the second biggest market in the world according to its gross domestic product (GDP) (International Monetary Fund, 2018), it is quite surprising that the role Europe plays in B2B sales is almost neglectable. It seems like Europe is on the verge of missing the trend of the early 21st century, while e-commerce is changing the working processes of the digitalized sector in the rest of the world (Terzi, 2011).

Taking a specific look at the chemical industry this picture does not change much. Being part of the 'old economy', changes do happen slower than in other industries and focus on other areas, respectively. Although, the number of investments is increasing in general (CEFIC, 2017), the chemical industry in Europe seems to lag one step behind internationally when it comes to marketing and sales (VCI and Deloitte, 2017). The industry still relies on old principles, i.e. business is generally conducted on a person-to-person basis. Suppliers or customers are acquired through extensive searches, contacted via phone and asked for quotes. Often, several phone calls and meetings are necessary to finally come to an agreement. The other way around works similarly: suppliers send their sales personnel to potential buyers to promote their goods and offers. Despite this extensive and time-consuming effort for both sides, the numbers of quotes buyers end up with as well as the conclusions of contracts the sales persons achieve often remain rather limited.

Once a contract is closed, it often results in a long-lasting cooperation, as the finding process is so complex as well as time-consuming. One could argue that long-lasting partnerships ensure stable businesses and should thus be beneficial for both sides. And, indeed, for the decades when telephone and fax were the most efficient communication channels, this is true. In times of globalization and interconnectedness, however, the benefits of personally knowing your supplier or customer cannot outweigh the advantages of modern technologies. Compared to the manifold possibilities opened up by digital technologies, the traditional processes appear rather slow, expensive and inefficient – not to say outdated. Also, suppliers are often using the opacity of the market for their benefit. The sales person visiting potential customers makes individual offers for each one, considering the size of the company, the insights the customer has on the pricing of certain products and so forth. The results are extensive negotiations about the conditions and prices.

3 Increasing number of marketplaces and web shops for chemicals

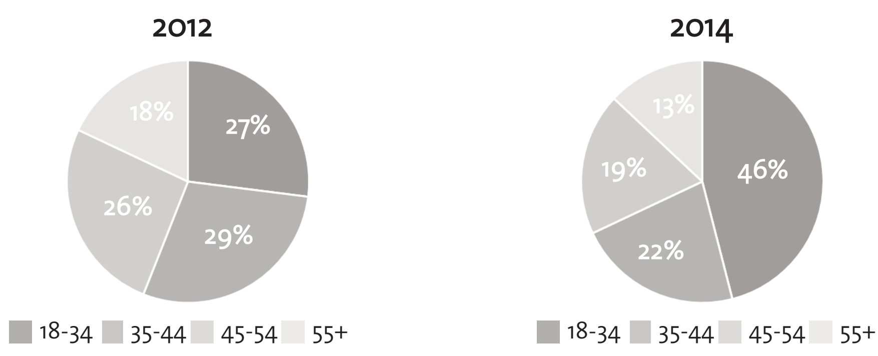
Looking at the number of marketplaces and web shops for chemicals on a global scale, there appears to be an abundance. So far, Europe only hosts a few promising but insignificant ones. Not all of the global marketplaces are specialized in chemicals only and they vary considerably in terms of their size, range and amount of offers. In order to get an overview of the various offers from different suppliers, commercial buyers of chemicals have to search, visit and compare multiple websites. Although, this process does save some time compared to the acquisition by phone, it is basically just a direct shift from the interaction on a person-to-person basis to a digital one. The information on products and offers is not spread personally but online, while buyers still have to search and organize it themselves. This might be somewhat more efficient, but it is still not close to what one would expect from digital services. The more marketplaces and web shops for laboratory and/or industrial supplies appear, the harder it gets to find the best offer. Hence, the development of the B2B e-commerce in the chemical sector is the direct transformation of the established principles and processes to digital platforms and communication channels. Compared to other sectors and even our private lives (e.g. travel websites, online shopping), it does not seem to be the optimal solution.

4 The European chemical industry is on the verge of being disrupted

While it was possible to rely on what worked for decades, the European chemical industry, which still is the third biggest in the world (CEFIC, 2017), now faces the risk of losing the lucrative sales side. What led those companies to be at the top once, ironically, is the very same thing enhancing the risk of being disrupted by newcomers now (Christensen, 1997). Over the past decades, many upcoming companies managed to outsmart incumbent players in other industries by making use of technological developments and/or radically rethinking business models (Markides, 2006; Yu and Hang, 2010). And first signs of a similar development for the chemical industry are already on the horizon when shifting the view to Asia or considering young and small players in Europe. Relative newcomers put forward innovation, introducing “a different set of features and performance attributes” (Govindarajan and Kopalle, 2006, p. 190). For the incumbent players those new features and attributes are hard to embrace for three main reasons:

1) **Listening to the best customers:** Companies are led in order to pursue sustaining the current successful state instead of aiming for disruptive innovation (Daneels, 2004). In the past main customers of chemical companies did not request digital solutions on the marketing and sales side. Thus, the companies were caught by them and shifted their activities to the customers promising the highest profit (Slater and Mohr, 2006). However, the (still) best customers become older eventually, and a young and technology-affine generation is constantly growing into positions, where they have the responsibility and the power to decide what to buy from whom (Snyder and Hilal, 2015). This young generation has been using e-commerce platforms for years in their private life and has no reason to make an exception for their business life. This generation of so-called millennials (aged between 18 and 34) is constantly striving for modernization to improve their business' efficiency, i.e. they want to achieve more in a shorter period of time compared to their predecessors. In business life, efficiency is considered the key for success and is rated higher

Figure 2 Share of people in B2B purchase positions by age groups (source: Snyder and Hilal, 2015).



than personal relationships with sub-contractors or suppliers. Young managers saw and experienced the rapid development of digital technologies during the past years and developed alongside it. More and more companies are embracing the possibilities the internet offers for their activities (Terzi, 2011). It has become normal to search for solutions to any problem online, as it is direct, fast and efficient. As shown in figure 2, the number of millennials in purchase positions has increased by 70 % from 2012 to 2014 (Snyder and Hilal, 2015) and there is no reason to assume that this trend has changed during the past years.

2) Deeply embedded routines and processes: Long-established principles are not suited for pursuing disruptive innovation and need to be adapted (Macher and Richman, 2004; Nelson and Winter, 1982). In this case, routines and processes, especially, refer to the ones applied for the evaluation of new initiatives (Assink, 2006). Disruptive projects often do not fulfill the company's expectations in terms of market size or return and are much more uncertain at first. If disruptive projects are assessed by the same criteria on a routinized basis, they might get sorted out prematurely (Christensen, 1997, 2006). Routines are capable of securing efficiency in a stable environment but show weaknesses during times of change due to inflexibility (Nelson and Winter, 1982). Therefore, relevant routines and deeply embedded processes can become obstacles decreasing the manager's ability in responding to disruptive innovations (Adner and Snow, 2010; Henderson, 2006). And if an initiative is not allocated with sufficient resources, it is prone to fail (Bergek et al., 2013). A study by KPMG (2016), conducted in Germany, reveals that in many companies, the staff is highly skeptical about new technologies. Afraid of investing in the wrong technologies or their employees not accepting the potential innovation, executives barely include budget for digital transformation of the company in their business plans. Without a budget, however, the number of qualified personnel able to evaluate the potential of new technologies and the benefits for the company remains low. Then again, the lack of adequately skilled employees to initiate and accelerate the digital progress of the company causes the skepticism about new technologies (KPMG, 2016).

3) Fear of cannibalization: This fear describes the effect of a firm's product directly diminishing the returns of another firm's product (Schilling, 2010). The willingness to execute projects possibly entailing this effect is described as the "extent to which a firm is prepared to reduce the actual or potential value of its investments" (Chandy and Tellis, 1998, p. 475). Accordingly, the introduction of new business models will be delayed, if company leaders are not prepared to reduce the value of their current business model and the fear of cannibalizing their own sales, assets or organizational routines prevails. Just like cost factors, this leads to a slow implementation of new business models. This was observed in the European chemical sector for a long time as the incumbent players rarely came up with own initiatives or undermined initiatives of other players successfully. By creating or supporting business models reducing the opacity of the market a beneficial aspect for the suppliers would falter. The organizational structures and the personnel are tuned to work in the old opaque model. Turning to a transparent model would render some investments obsolete by cannibalizing own sales.

Due to those inhibitors only 30 % of the medium-sized companies in Germany have yet implemented digital business models, according to survey studies by VCI and Deloitte (2017). The remaining 70 % of the participants indicated that they only expect minor changes to their business models due to the ongoing digital transformation. Although, new distribution channels are often mentioned as important factors for improvement. In a recent study, Euroforum asked 50 experts from the chemical sector about the digital transformation of chemical companies. The survey revealed that most of the participants see the most important field for improvements by digitalization in the supply chain management not in new business models (Euroforum, 2017). Accordingly, digital distribution models are implemented hesitantly. In addition to the general skepticism, the additional expenses for the creation and maintenance of a web shop (which also requires an online marketing strategy to effectively increase the range of the company) are the main reasons for the hesitation. Many companies lack the personnel resources for such projects. However, in order to comply with the expectations of potential customers, online presence has already become much more than just an extra service. It is a necessity for professional businesses. They can answer customer inquiries in real-time, adapt new online offers, and analyze traffic and range. This way, businesses can make use of the digitalization to check their sales models, achieve predictable results, and generate long-lasting customer relationships, despite the growing number of competitors. Online marketplaces and web shops become increasingly relevant for many sectors (Euroforum, 2017).

5 An agile startup without the incumbent's constraints presents an all-in-one-platform solution

For years, the chemical industry mainly consisted of established players, who, for a long time, cemented their position on the market and did not seem to leave room for smaller startups. With the upheaval into a new, digital era, this is due to change. An increasing number of startups strive to innovate the chemical sector and help leading it into the times of 'marketing and sales 4.0'. This also applies for the first global metasearch engine for chemicals and related services. A startup company from northern Germany is the first to introduce such a concept to the sector.

The founders of chembid know the problems they want to tackle first-hand. Dealing with the opaque and complex market situation every day, the idea for a more efficient solution suited for the new digital era arose. In 2016 the company was officially founded. It took one and a half years to design and develop the metasearch engine and platform of the same name, which went online in October 2017. The idea and the aim is to collect, organize, and bundle commercial and technical information on chemicals from all over the world into one website, to make it easily accessible and guarantee simple and fast processing. In contrast to the incumbents chembid is not limited by its past. Hence, the company is not constrained by static routines or a fear of cannibalization and the team can strive to develop a disruptive innovation for the future customers (Weiblen and Chesborough, 2015). Following this target would have been hard to accomplish staying along the beaten track.

5.1 The first metasearch engine for chemicals

The backbone of the platform is a metasearch engine. And although the technology itself is not new, it is new to the chemical industry and was adjusted to fit the demands of it. The platform can be found and visited like any other website on the internet. Traffic is generated from generalist online search engines, social media networks or other websites (e.g. via online marketing campaigns), etc. Being a highly specialized search engine itself, it can schematically be placed one level above the respective marketplaces for chemicals and the suppliers' individual web shops, if available. The data is gathered in two distinct ways. On the one hand the relevant websites are searched for suitable offers by the intelligent search engine. So-called web crawlers are used to automatically search for product information from thousands of offers worldwide. On the other hand the platform uses application programming interfaces (API), which allow for the simple linking with other databases. Via those APIs, providers of marketplaces or web shops can easily, i.e. manually in a few steps or even automatically, transfer their offers onto the meta platform. The information is then made available in a single overview and can be narrowed down with chemical-specific filter functions.

On this meta-level of the search engine, the platform is not a direct competitor for online marketplaces for chemicals, but simply bundles the information from various competitive marketplaces on one website. Partnerships with marketplaces and suppliers with their own web shops are even desired, as it ensures the accuracy of the information and offers. Being listed on the platform does not take away traffic, as all offers are linked to the original website, where purchases can be carried out as usual. However, the platform does offer its own marketplace for further opportunities as well. The search engine does not prioritize offers from the inbuilt marketplace, though, as this would disrupt its main purpose of creating more market transparency.

In addition to the product database a supplier database is included as well and can be browsed, too. The suppliers are linked to their respective chemicals and related services. This data is used to create a comprehensive worldwide data-base of supplier profiles and their products and online offers. Besides generating a value for itself as being the first index to link suppliers with their actual online offers, it is a highly important building block for the sourcing tool, which is another innovative feature of the platform.

5.2 Sourcing tool – reversing the market

The online sourcing tool for chemicals and related services can help both vendors and buyers to increase their efficiency. Buyers can place their demands, wait for the responses of suppliers and compare the quotes in a standardized overview to choose the best offer. Vendors on the other hand can save on sales force, who would usually contact every potential buyer. They can send quotes fitted to the buyer's request. This is somewhat reversing the market situation, but it also helps create a more balanced distribution of power. So far, so old. The unique point is derived from the combination of the gathered supplier and product data and an algorithm finding suitable suppliers for the product in question. The algorithm will check for suppliers fulfilling the specific criteria of the request. For example, typical amounts of the suppliers are compared to the amount requested. Accomplishing a more and more accurate algorithm will be a big step towards connecting buyers with potentially interested sellers only fully automatically.

5.3 Market statistics - big data, small effort

Since the process of providing such services includes the collection of a vast amount of data, it is used to create useful market statistics as an extra service for the users. While it was a time-consuming process to get reliable price information due to the lack of market transparency in the past, the data gathered by the web crawlers allow for the reliable computation of past prices, average prices or future trends. Based on this data, the statistical tool offers the possibility to create clear tables and graphs according to the users' desired criteria. This automatically delivers a representative result from a big pool of suppliers rather than from a few who need to be called manually. The possibility to handle big data for the chemical B2B sales sector is made easily accessible for the users.

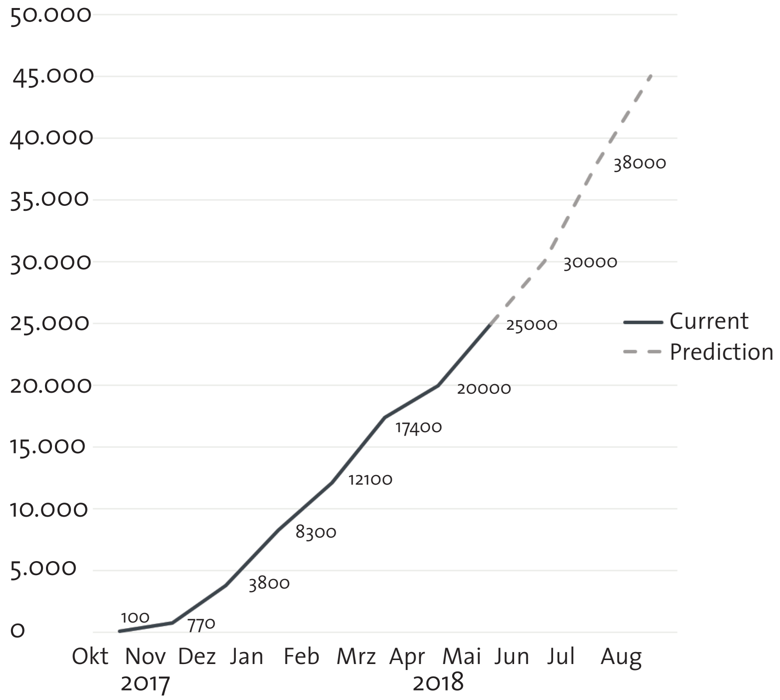
5.4 Artificial intelligence - virtual agent

In order to guarantee the best search results possible, Artificial Intelligence (AI) plays a significant role in the form of machine learning algorithms monitoring the users' search behavior and adapting the search results according to their needs and preferences. Due to the abundance of offers, this saves users additional time, as it helps to find products for certain cases. At first, this feature can help rudimentary, but over time, with more and more data gathered, the AI will be able to advise in more advanced cases and serve as a virtual agent. As one of the main goals of the platform is to increase efficiency, the usage of AI/machine learning is vital to accelerate the processes while still providing qualitative results.

5.5 Advantages for buyers and suppliers

As a result of the ongoing digitalization and globalization in all sectors, vendors are put under increasing pressure (Matthyssens et al., 2008). They need to cope with a growing number of competitors from all around the globe as well as the high expectations of their customers. To stay competitive, they need to strive for more coordination and collaboration among supply chain partners to increase efficiencies in supply chain management (Terzi, 2011). An online platform solution, offering new distribution channels for sellers, can help firms to easily expand their range without requiring additional resources for e-commerce and online marketing (provided they chose a widely renowned marketplace). The unique all-in-one solution from Germany, for instance, had approx. 25,000 unique visitors in May 2018, while the projected number

Figure 3 User development on chembid (source: own representation based on Google analytics).



of users for August 2018 is estimated to be 45,000. Due to the usage of cloud computing the model is easily scalable and there is no hard limit to those numbers. The platform can thus be a useful addition to existing marketing and sales solutions. New leads are automatically acquired, so that marketing resources and sales force can be reduced. Additionally, the marketplace can serve as an alternative to an individual web shop or as an addition to it.

After creating a user account and being verified by the provider, product or service offers can either manually or automatically be uploaded onto the inbuilt marketplace. Big amounts of data, e.g. a supplier's whole range of products, can be uploaded in bulks. All offers can be found on the metasearch engine. Offers from the inbuilt marketplace are highlighted, for better findability of verified vendors. The number of current available offers, i.e. 1.8 million, and the 60,000 represented suppliers from more than 160 different countries, prove the platform's potential.

The advantages of a global metasearch engine for commercial buyers of chemicals are quite straightforward as well: Buyers are usually the ones actively seeking for information and the best offer. Upon entering the name or CAS number of a product into the search bar, they will receive a comprehensive list of suitable offers from online marketplaces and websites worldwide. Various filtering options, e.g. place, price, quality and measurement unit, allow for further specification of the search request. Results can be sorted by price or relevance. Clicking on the search result will directly lead to the supplier's website or the marketplace the offer can be found on.

6 Summary and conclusion

- Digitalization reaches out to the B2B marketing and sales sector and makes no exception for the chemical industry. European chemical companies, though, do not pursue sufficient digital initiatives. By going on conducting business in the old ways and without preparing for what is to come, other players – especially from Asia – might pace them out.
- Asian companies already developed an abundance of marketplaces and offer chemical products online. The growing number of those marketplaces leads to a mere transformation of the old, time-consuming and inefficient processes to digital channels.
- Three main factors are limiting incumbent European players in starting their own initiatives seriously: Listening to the best customers, deeply embedded routines and processes and fear of cannibalization. Hence, only 30% have implemented digital business models so far.
- As a young startup, chembid is not limited by the incumbent's constraints and is, thus, a good example for a potentially disrupting innovation for the B2B sales sector of the chemical industry. The company developed the first metasearch engine for chemicals and related services. Its main purpose is to create full market transparency by organizing and making accessible chemical offers from the web.
- Besides other advantages this approach serves suppliers with an increased range and buyers with more and better market information.
- Among other features the holistic 'chembid concept' also includes market statistics based on big data techniques and a sourcing tool run by AI. By combining modern technologies, it is aimed to co-create the future of chemical businesses regarding marketing and sales.

References

Adner, R., Snow, D. (2010): Bold Retreat, *Harvard Business Review*, **88** (3), pp. 76–81.

Assink, M. (2006): Inhibitors of disruptive innovation capability: A conceptual model, *European Journal of Innovation Management*, **9** (2), pp. 215–233.

Bergek, A., Berggren, C., Magnusson, T., Hobday, M. (2013): Technological dis-continuities and the challenge for incumbent firms: Destruction, disruption or creative accumulation?, *Research Policy*, **42** (6-7), pp. 1210–1224.

CEFIC (2017): *Cefic - Facts and Figures 2017*, available at http://fr.zone-secure.net/13451/451623/#page=1, accessed 24 May 2018.

Chandy, R. K., Tellis, G. J. (1998): Organizing for Radical Product Innovation: The Overlooked Role of Willingness to Cannibalize, *Journal of Marketing Research*, **35** (4), pp. 474.

Christensen, C. M. (1997): *The innovator's dilemma:When new technologies cause great firms to fail. The management of innovation and change series*, 2nd ed., Harvard Business School Press, Boston, Mass.

Christensen, C. M. (2006): The Ongoing Process of Building a Theory of Disrup-tion, *Journal of Product Innovation Management*, **23** (1), pp. 39–55.

Danneels, E. (2004): Disruptive Technology Reconsidered: A Critique and Re-search Agenda, *Journal of Product Innovation Management*, **21** (4), pp. 246–258.

Euroforum, (2017): *Industrie 4.0, Index 2018 – Der Wegweiser in die digitale Zu-kunft*, available at: http://www.euroforum.ch/strategietagung/

download-digitales-magazin-industrie-4-0-index-2018-der-wegweiser-in-die-digitale-zukunft-pc/ accessed 20 May 2018.

Govindarajan, V., Kopalle, P. K. (2006): Disruptiveness of innovations: Meas-urement and an assessment of reliability and validity, *Strategic Manage-ment Journal,* **27** (2), pp. 189–199.

Henderson, R. (2006): The Innovator's Dilemma asa Problem of Organizational Competence, *Journal of Product Innovation Management*, **23** (1), pp. 5–11.

International Monetary Fund (2018): *World Economic Outlook Database*, available at http://www.imf.org/external/pubs/ft/weo/2018/01/weodata/weorept.aspx?sy=2017&ey=2017&scsm=1&ssd=1&sort=country&ds=.&br=1&pr1.x=45&pr1.y=10&c=946%2C137%2C193%2C181%2C124%2C918%2C138%2C964%2C182%2C960%2C423%2C968%2C935%2C128%2C939%2C936%2C961%2C172%2C132%2C184%2C134%2C174%2C144%2C944%2C178%2C136%2C112%2C111%2C941&s=NGDPD&grp=0&a=, accessed 24 May 2018.

KPMG (2016): *Zeit zum Aufblühen – Digitalen Transformation der chemischen In-dustrie*, available at: https://home.kpmg.com/de/de/home/themen/2016/11/digitale-transformation-der-chemischen-industrie.html accessed 24 May 2018.

Macher, J. T., Richman, B. D. (2004): Organisational responses to Discontinu-ous Innovation: A Case Study Approach, *International Journal of Innovation Management*, **8** (1), pp. 87–114.

Matthyssens, P., Kirca, A., & Pace, S. (2008): Business-to-business marketing and globalization: two of a kind, *International Marketing Review*, **25** (5), pp. 481-486.

Markides, C. (2006): Disruptive Innovation: In Needof Better Theory, *Journal of Product Innovation Management*, **23** (1), p. 19–25.

Nelson, R. R., Winter, S. G. (1982): *An evolutionarytheory of economic change.* Cambridge, Mass.: Belknap Press of Harvard University Press.

Samiee, S. (2008): "Global marketing effectiveness via alliances and electronic commerce in business-to-business markets", *Industrial Marketing Management*, **37** (1), pp. 3-8.

Slater, S. F., Mohr, J. J. (2006): Successful Development and Commercializa-tion of Technological Innovation: Insights Based on Strategy Type, *Journal of Product Innovation Management*, **23** (1), pp. 26–33.

Snyder and Hilal (2015): T*he Changing Face of B2B Marketing*, available at https://www.thinkwithgoogle.com/consumer-insights/the-changing-face-b2b-marketing/, accessed 24 May 2018.

Statista (2017): *B2B e-Commerce 2017*, available athttps://www.statista.com/study/44442/statista-report-b2b-e-commerce/, accessed 24 May 2018.

Terzi, N. (2011): The impact of e-commerce on international trade and employ-ment, Procedia - *Social and Behavioral Sciences*, **24** , pp. 745-753.

VCI and Deloitte (2017): *Chemie 4.0 – Wachstum durchInnovation in einer Welt im Umbruch*, available at https://www.vci.de/vci/downloads-vci/publikation/vci-deloitte-studie-chemie-4-punkt-0-langfassung.pdf, accessed 24 May 2018.

Weiblen, T., and Chesbrough, H. W. (2015): Engagingwith startups to enhance cor-porate innovation, *California Management Review*, **57** (2), pp. 66-90.

Yu, D., Hang, C. C. (2010): A Reflective Review ofDisruptive Innovation Theo-ry, *International Journal of Management Reviews*, **12** (4), pp. 435–452.

Practitioner's Section

Successful data applications: a cross-industry approach for conceptual planning

Robert Jenke*

* Dr. Jenke Consulting, Process Automation & Data Science, www.jenke-consulting.de, mail@jenke-consulting.de

DOI: 10.17879/38129709602; URN: nbn:de:hbz:6-38129709797

Data-based solutions powered by artificial intelligence (AI), and especially its subdomain of machine learning, are a key driver of today's fast-paced technological evolution. In the process industry, barriers for many organizations to apply this technology are missing know-how for conceptual planning as well as lack of economic feasibility studies. However, companies risk to lose their competitive position by not applying this technology. In this article, we describe the CRISP-DM model as a conceptual planning approach. In addition, we provide practical advice based on experience in other industries how this technology can be applied in the process industry. Here, the steps process analysis and data understanding are key success factors in order to develop economically viable use cases. An implementation strategy should include an agile environment to develop ideas fast and with little risk before transferring working solutions to the requirements of the operational business. A combined bottom-up / top-down approach of knowledge distribution and pilot projects can help organizations to successfully embrace this technology in their operational businesses, overcome associated fears and organically seize company-individual opportunities that arise.

1 Introduction

The fast-paced technological evolution of our time leads to equally fast changes in the business world, from efficiency gains through process improvements to the redefinition of long-established business models. In this context, a key technology, which accelerates this technological evolution, is artificial intelligence (AI) including its subdomain of machine learning. Machine Learning methods allow the training of models from data examples instead of tediously having to define the input-output relationship of the models manually. For instance, such models can be applied to make decisions and predictions. While this technology has been cost- and time-consuming in the past, it has now reached a level of maturity where broad industrial application is feasible in many scenarios (VanThienen, 2016).

As AI constitutes a universal technology – like the steam engine or electricity in the past – it is bound to influence many sectors of industry well beyond its origin (Brynjolfsson, 2017). Consequently, also the process industry is already being influenced by this trend and therefore evaluates how to harness the arising opportunities. For example, predictive asset management has been adopted by major players in order to maximize asset utilization or minimize unplanned downtime. Furthermore, the monitoring of a plant can be made more efficient by the digitization of a plant's control data. Additionally, monitoring of production processes can be enhanced with pattern recognition in order to assess influences on batch consistency or to predict deviations from the production process before they could occur. Finally, downstream data from points-of-sale can be used to forecast the demand of customers and to plan a responsive schedule in production (VanThienen, 2016).

However, at the same time several barriers are currently hampering the use of digital technologies like machine learning, as Stoffels and Ziemer point out in an article of an earlier issue of this journal (Stoffels and Ziemer, 2017): Among those barriers are

1) Unclear benefits and/or lack of economic evaluations and
2) Missing know-how on methods for analyzing and adapting processes.

Yet, it is imperative for companies to understand the impact that this important technological change has on their business model, their operations and their competitive landscape. As with any new technology, identifying and defining use cases is not straight forward. At the same time, they need to fulfill at least two preconditions:

1) the application must be technologically feasible and
2) backed by a solid business case.

If these preconditions are not fulfilled in the beginning, many companies eschew investments due to risks involved. However, there is also considerable risk associated with ignoring the changes ahead and thereby risk becoming the next Kodak (Anthony, 2016).

In this article, we aim to make the following contributions:

1) We present methods and approaches concerning data applications that have been successfully applied to other industries. More importantly, this contribution shows what can be learned and transferred to the process industry. In this, we introduce the cross-industry standard process for data mining (CRISP-DM) model, which is a generic framework how to approach data mining problems. We discuss key issues in depth based on experience from other industries.
2) We discuss possibilities to overcome the previously mentioned risks and barriers, i.e. implementation strategy and integration into the organization on a strategic level, in order to increase the overall ROI of investments in this field.

2 CRISP-DM model

The CRISP-DM describes a generic framework for data mining and knowledge discovery projects. It is used by the majority of experts in the field of Data Science (Marban, 2009). The approach captures the essence of what is needed to successfully carry out a data mining project, but at the same time it is easily transferable to any industry. The six steps of the CRISP-DM are Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Shearer, 2000). It is worth noting that the steps are not necessarily carried out in a linear manner, as depicted in Figure 1. The single steps are explained below (for details, see (Shearer, 2000)):

1. Business understanding
Before touching any data, it is very important to understand the business context in which this data will be used. This phase concerns itself with questions around the business objectives, success criteria, relevant business processes, and an assessment of the overall situation, in order to derive goals and develop a first plan of the project. The better this step is carried out, the more valuable insights and outcome can be expected from the remaining steps.

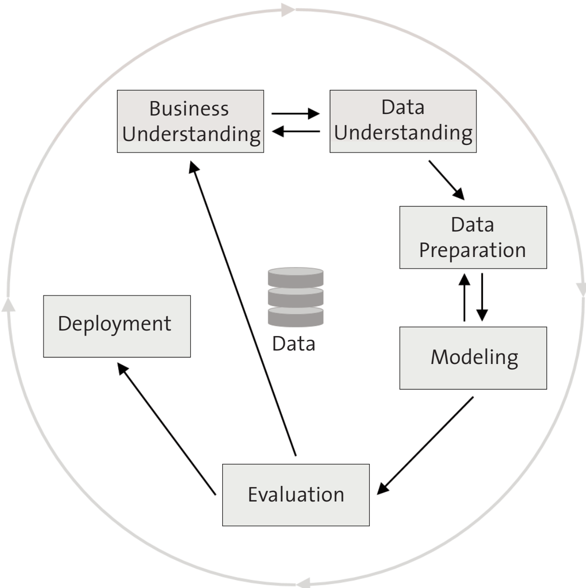
2. Data understanding
In the second step, data is collected and an understanding is developed. This includes describing the data and what kind of information it contains. Also, quality of data should be assessed, e.g. consistency and completeness. First hypothesis can already be derived from data. Often, it is necessary to go back to step one multiple times, until a clear picture of the interaction between business processes, objectives, and data is gained.

3. Data preparation
Once a specific use case is defined, data must be selected based on the relevance to the project's goal as well as the technical limitations. Here, the data must be usable. Data sets must be provided, which can be used to develop models. Quality can be increased, for example, by removing outliers, handling missing values, and giving a proper structure to the data.

4. Modeling
In this phase, tools like machine learning methods are employed to train one or several models on the selected data. Usually, more than one method is available. This phase aims to identify the best fit and optimize any free parameters. This also includes testing the quality of the model, e.g. its generalizability measured by the error rate on test data.

5. Evaluation
While the model itself is verified in the fourth step, this step evaluates the suitability of the developed processing pipeline and model with respect to business application. Only when there are no critical issues overlooked, the model can be deployed in the next step. Too often, false assumptions require

Figure 1 CRISP-DM model (source: Shearer, 2000).



going back to step 1 and to revise the business understanding.

6. Deployment

When the suitability for real-life business application has been shown, the deployment phase aims to transfer the findings from the data mining project to day-to-day operations, i.e. actually make use of the created models. In some cases, this might be a simple report, in others the implementation of software to track and analyze real-time data in order to support the decision-making process of the organization.

In Data Mining more than anywhere else, a famous quote, that may or may not be attributed to Albert Einstein, can be literally applied for good results: "If I had one hour to save the world, I would spend 55 minutes understanding the problem and 5 minutes trying to find a solution." Challenges frequently arise in the definition of use cases, i.e. truly understanding the problem. Therefore, understanding business and understanding data need to be carried out thoroughly and will have to go back and forth. We collected learnings from experience gathered in other industries and summarize these in the following together with some practical examples.

2.1 Process modeling and analysis

The step of understanding business is very much about understanding processes, i.e. identify tasks and steps that data can improve, enhance or carry out more efficiently and reliably. Therefore, modeling tools like Business Process Modeling Notation 2.0 (BPMN 2.0 (OMG, 2013)) are helpful to model processes and develop a common understanding. Although the BPMN 2.0 notation consists of a large number of available elements, a handful of the basic elements are already sufficient to significantly increase process understanding.
Approaches to generate use cases can be versatile. A high-level approach is to focus on areas of the value chain that have an impact on operating and growing the business, as van Thienen et al. suggest (VanThienen, 2016). On a more practical level, it should be looked out for the following cues to identify high-potential areas for data and/or AI usage when analyzing processes:

1. High-volume tasks:

Naturally, task that are frequently carried out often provide a greater leverage for potential improvements. In these cases, even small time saving measures or quality improvements can lead to a viable business case for applying machine learning.

2. High-value decisions:

Similarly, decisions with a high inherent potential to influence large areas of the organization are promising candidates for intensive data-usage. In these cases, additional efforts may be well justified to make the right business decision.

3. **Media discontinuity**:
Despite increasing efforts to digitize, many processes still involve media discontinuities, e.g. printing of documents or form, switching from online workflows to phone calls or emails. While sometimes this is obviously not avoidable, e.g. sending a hardware product to the end-customer, steps within the process that go back and forth between digital and analog workflows are good candidates for applying automation.

4. **Time consuming steps:**
A strength of AI methods is to process large amounts of data fast without any errors. This often complements human abilities and can lead to significant time savings. Thus, it is worth focusing on the annoying and tedious tasks.

5. **Bottlenecks:**
Restructuring processes around bottlenecks can increase overall efficiency. A solution can be parallelization of sub-tasks carried out simultaneously by a machine in the background.

As with any new product or service, it is important to take on a "Customer centric" mind-set and develop the solution this way. Here, the customer could also be an internal one. In the following, we like to illustrate this approach with an example from practice.

Example: In one project, we analyzed the internal processes of the sales department. One of the planning steps consumed half a day on average and was carried out regularly. The organization was growing substantially, so that the process needed to be rethought for scalability. In a workshop, we modeled the current state of the process and analyzed possibilities for restructuring including the automated processing of data. We were able to eliminate all media discontinuities by fully digitizing the planning process through a web-application. Large parts of the data were then gathered and processed automatically. Overall, this led to significant time savings of up to 80% and further increased quality and transparency of this process.

2.2 Data understanding and preparation

The second important step is to understand and prepare the available data. Both technical and economic feasibility of a use case depend strongly on the data available, which makes use cases highly individual according to a company's data situation. Therefore, several topics need to be taken into account when aiming to use data successfully. Put in simple terms, the following factors influence the potential of your data and how well you can build models from it. They can be considered as levels of data needs building on each other. Each level requires a certain maturity of the previous level:

1. **Data sources and types**
First, where does your data come from and since when are you recording it? Is it a structured SQL-Database, a sensor-stream, a manually filled spreadsheet-file, or something else? Data can be structured or unstructured. For example, a free text document like a project report or an invoice in pdf-format is considered unstructured data. If you have a form like a contact form on a website, then each field (e.g. name or email address) has a particular meaning. This represents structured data. Unstructured data is much harder to process and requires more steps to prepare the data.
Data types regard to the kind of data. For example, this can be text, numerical, or categorical. The latter refers to a fixed set of values that the data can take on, e.g. 'True' or 'False' or colors of a product 'red', 'green', or 'blue'.

2. **Data quality**
Second, it is important to assess and understand the quality of the data available. In particular, the following questions are of importance: Are the values complete over the whole data set or are there missing values? How should they be handled and how does this affect the quality of the application? Can the root problem be fixed or can the data be processed even when values are missing? Are outliners present in the data? If so, the data might need to be cleansed. Many of these rather technical issues can also point to issues in the process setup itself.

3. **Data content**
Third, it is important to understand what information does this data represent? With numerical data, descriptive statistics can be used to describe the data (i.e. data distribution, histograms, etc.). Although this step may seem tedious and without a clear goal at first, every minute spent is well invested. It is a prerequisite to understand the data available very well. Then, a judgment of the models concerning their suitability and application is possible at later stage of the process. When digging into the data, ideas for potential use cases can be generated. In most cases, insights about operational and organizational processes can be gained, which are not expected at first.

Example: At one of our customers, production controlling data was recorded in both Enterprise Resource Planning (ERP) system and Manufacturing Execution System (MES), but was not intensively utilized. Before creating use cases to apply AI, the available data was inspected for its types, quality and content. This exercise proved to be highly educational for everyone involved. In fact, untapped information was identified, and a new detailed report personalized for each foreman was designed and introduced as a result. This example illustrates, that effective solutions to enhance data usage can also be quite simple.

2.3 Modeling and evaluation

In the data modeling step, the actual building of models using machine learning methods takes place. Models are a representation of relationships. As shown in Figure 2, they map input data to an output. In the past, models were often created manually. For example, if-than-that-rules can make up the relationship between input and output. The strength of machine learning methods is that they can extract relevant information like rules from data automatically. By doing this, the model can be learned from existing data, i.e. input-output examples from the past. Against common belief, this step consumes only a relatively small part of the time – thanks to the versatile tools like software packages available today. These tools facilitate the implementation of machine learning methods to train and evaluate models.
Machine learning methods can be summarized in three categories:

1) Supervised Learning, where the model output is provided to the machine learning method for training. For example, Classification, Regression, and Feature Subset Selection methods belong to this category.
2) Unsupervised Learning, where the model output is unknown. That means, the methods have to find structure and make connections by themselves. Clustering and Dimensionality Reduction are examples here.
3) Reinforcement Learning, where the model is trained independent of the output. Instead a trained rule is reinforced by rewarding / punishing depending on whether the output was right / wrong. Artificial Neural Networks and Deep Learning can be assigned to this category.

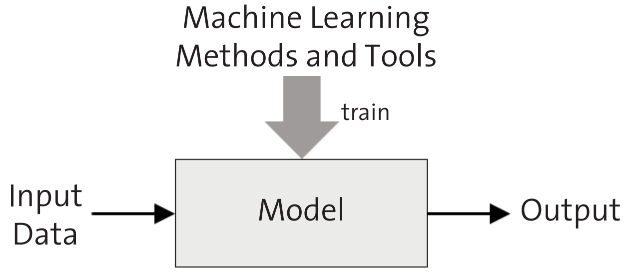
For a given problem, the model can usually be created by means of several methods. The challenge for data scientists is to pick the appropriated method (or a combination of methods) based on their experience in order to develop the best model. Models are trained based on examples. Therefore, it is important to consider the generalization of the model very well, so that unseen examples are also handled well by the model. Thus, the quality of the model is judged based on a so-called “test-set” of data that is put aside in the beginning of the modeling phase. The following example illustrates how pattern recognition can help to improve processes of a plant.

Example: A plant’s control data from the past contains information about process deviations and when they occurred. Pattern recognition can be applied to automatically identify such deviations before they occur. This task can be framed as a classification problem with two classes: “deviation” or “no deviation”. From past examples of plant operations, relevant data features and rules to detect such deviations are extracted using machine learning methods from supervised learning. Before the final model containing these rules is deployed, its fit is evaluated using a second set of examples from the past.

2.4 Implementation strategy

The understanding of both process and data provides two important results. First, the technical feasibility as well as the potential benefits of a use case can be judged. Second, the effort how to implement the respective use case can be roughly estimated. This provides an indicator of economic fea-

Figure 2 Machine learning methods allow to train models from data examples (source: own representation).



sibility, which significantly lowers the risk of investing disproportionate budgets. However, a considerable amount of ideas might still be disregarded at a later stage due to factors that cannot be assessed at this point. In such cases several key factors help to reach a good ROI. In particular, methods can be used that are already widely applied in software development and the lean start-up approach, which focus on agile and iterative developments (Ries, 2011). Risks of failure can be reduced by allowing multiple possible directions. Then, adjustments can be made based on the learning made within the implementation process. Organizations should start with prototypes at a very early stage to generate this learning. Tackle assumptions with highest risks first, i.e. fail fast to learn fast! This, of course, requires a culture that is open-minded towards failure.

Meanwhile, it is important to keep a clear focus. A strong and rigorous selection funnel is necessary to weed out directions, which are not promising. Only projects should be continued that have proved to be successful by fulfilling the defined evaluation criteria.

The CRISP-DM model aligns very well with this lean approach. This approach will be illustrated by the example below.

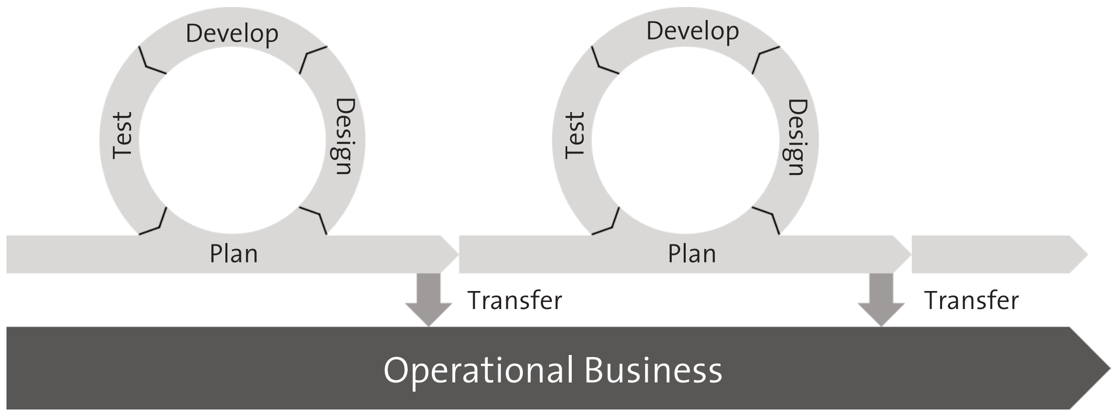
Example: One company holds monthly meetings, in which employees participate that are not involved in the actual project – similar to a "supervisory board". Learning from the project regarding what works and what does not work as well as new ideas are presented to them. The group then judges the value and feasibility of the projects and takes a decision to terminate directions that do not show enough potential with respect to the company's competitive position.

At first, these elements may seem difficult or sometimes impossible to incorporate in the development processes of companies from the process industry, which are traditionally rather conservative and have strict requirements in their quality systems. A method to solve this is depicted in Figure 3 – similar to Google's so-called moonshots[1]: This should be a parallel and isolated track to the operational business. In doing this, it provides a way to develop solutions to a level of maturity without a costly overhead and operational risks. Only when a solution has proven to provide benefits it is transferred to the operational business. This approach greatly decreases costs and risks associated with such endeavors. If a company decides to not establish such a department internally because of its size or budget, an outsourcing of such services to external experts can be considered.

2.5 Integration into organization

For the long-term success, it is vital to find a sustainable way as an organization to embrace the AI technology and to embed it into the organization's strategy. The subsequent adaptation of business processes also brings about change to the employees, which play an important role concerning the integration of such technologies. The human factor and the topic of change are easily overlooked. Thus, human factors such as their behavior must be especially considered. Several fears are associated with AI as the public discourse discloses. For example, its impact on job security and job descriptions. These fears must also be addressed within

Figure 3 Development of innovative applications in a parallel environment (source: own representation).



1 Moonshot refers to projects that are similarly ambitious and unrealistic as landing on the moon seemed back in the early 1960's, but that would constitute a significant advance.

the organization. One way to move forward is to follow a combined bottom-up and top-down approach:

Bottom-up: A key factor to get support from the workforce for technological change is the distribution of facts and knowledge. Employees should be encouraged to identify use cases by themselves since they know their workflows and tasks better than anyone else. This includes the exhausting and repetitive processes, which could be promising candidates for automation. The ideas of employees become a valuable source for process improvement, when they are equipped with the right knowledge about the technological capabilities and limitations. At the same time, they can be mobilized for the change and a momentum can be created, if they are included at the beginning of the transformation process.

Top-down: In parallel, pilot projects initiated by the management serve as vivid examples in a familiar environment help to spread a realistic and clear picture of what the organization's individual needs are and showcase success stories as a generator for new ideas.

3 Summary and conclusions

Advanced and complex technologies like AI require knowledge and a well-structured approach for conceptual planning and implementation. With the CRISP-DM model, an established framework is available, which can be used to guide projects that aim at the utilization of data in order to get valuable insights. In this process, challenges frequently arise regarding the definition of use cases. In this article, insights from practical experience offer approaches how to overcome these challenges.
For the conduction of projects, modern and agile development methods such as the lean start-up approach should be used. Strict evaluation criteria concerning the decision which projects should be further pursued must be in place. Implemented correctly, this will ensure a high ROI across these efforts. Further, a holistic strategy for integrating this change into the organization is necessary to seize opportunities that arise organically.

Many companies are still reluctant to tackle the technological shift towards intensified data usage and artificial intelligence – they remain in a waiting position. Yet, it has to be considered whether the energy and resources spent in observing the market might not be better invested in gaining first-hand experience by diving into the topic. The first option leads to an inevitable time delay. Thus, the second option should be more appealing to companies. Especially, since both feasibility and economic viability are two company-individual factors. First practical experiences can be collected by just diving into the topic and testing assumptions.

References

Anthony, S. (2016): *Kodak's Downfall wasn't about Technology*. Harvard Business Review, available at https://hbr.org/2016/07/kodaks-downfall-wasnt-about-technology, accessed 10 July 2016.

Brynjolfsson, E., McAfee, A. (2017): The Business of Artificial Intelligence, *Harvard Business Manager*, pp. 22-34.

Marbán, Ó., Mariscal, G., Segovia, J. (2016): A Data Mining & Knowledge Discovery Process Model, in Julia, P., Adem, K. (ed), *Data Mining and Knowledge Discovery in Real Life Applications*, pp. 438-453.

Shearer, C.(2000): The CRISP-DM Model: The New Blueprint for Data Mining, *Journal of Data Warehousing*, vol. **5** (4), pp. 13-22.

Stoffels, M., Ziemer, C. (2017): Digitalization in the process industries – Evidence from the German water industry, Journal of Business Chemistry, vol **14** (3), pp. 94-105.

Object Management Group (2013): *Business Process Model and Notation BPMN 2.0*, ISO/IEC 19510.

Ries, E. (2011): *Creating the lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, Inc Magazine, available at https://www.inc.com/magazine/201110eric-ries-usability-testing-product-development.html, accessed 15 October 2011.

van Thienen, S., Clinton, A., Mahto, M.,Sniderman, B. (2016): *Industry 4.0 and the chemicals industry: Catalyzing transformation through operations improvement and business growth*, Deloitte University Press.

Practitioner's Section

Quantitative biology – a perspective for the life sciences' way into the future

Holger Wallmeier*

* CONDOR scientific computing & consulting, Sossenheimer Weg 13, 65843 Sulzbach/Ts., Germany
holger.wallmeier@condor-scientific.com

DOI: 10.17879/38129708916; URN: nbn:de:hbz:6-38129709089

Life science research and life sciences' industries are facing an overwhelming complexity of biology. Today's scientific methods and technologies allow for a very detailed look at biology. What is left to do, is understanding and interpretation. Quantitative biology, the close coupling of life sciences, mathematics, and statistics is likely to provide the methodologies to turn collected data into dedicated information and knowledge. The most promising approach is the formulation of mathematical models on the basis of machine learning. The predictive power of such an approach is a promising option for basic biological research, medicine, pharmacology, agricultural science, and ecology. Furthermore, also R&D of related life sciences industries can take advantage of this digital approach to meet future challenges and market requirements. Quantitative biology plays the role of an enabling technology.

1 Introduction: biology as a quantitative science

It is clear that today biology is influencing human thinking, perception, and action in an increasing number of different ways. The 21st century is seen as the century of biology (Venter and Cohen, 2004). Life sciences' historical route from genetics to genomics, now approaching and establishing synthetic biology shows its impact, not only on biological basic research, but also on the different disciplines of biotechnology, medicine, pharmacology, and last but not least, agricultural science. Consequently, there is a corresponding influence on the life science industry, including pharmaceutical industry, diagnostic industry, medical product industry, food as well as dietary supplement industry, and agricultural industry.

In a more general view the combination of biology, statistics, and mathematics has been termed quantitative biology (Zhang, 2013). The notion of quantitative biology may sound a little bit unfamiliar today. We are rather used to talk and hear about bioinformatics, computational biology, biometry, biostatistics, biomathematics, and similar disciplines, all of which share the combination of biology with some other scientific discipline, which is related to calculation, modeling, and computing. Though in the past, biology itself has been recognized as a predominantly descriptive science (Mayer, 1997), the roots of a quantitative view are fairly old, as can be seen in the proceedings of the first Cold Spring Harbor Symposium on Quantitative Biology, which Reginald Harris organized in 1933 (Witkomski, 2018). This may be seen as a reaction to the early 20th century discussion about biology as an autonomous science, or just a sub-discipline of physics and chemistry (Mayer, 1997).

Today, biology has established as a science of information, driven by molecular biology, genetics, and genomics. Functional genomics, as well as metabolomics have produced data that demonstrate the existence and importance of complex pathways and networks in living cells and organisms. The complexity of biological systems appears to be much higher than in today's technological implementations, which has become evident in applying, e.g. non-equilibrium thermodynamics, synergetics, and chaos theory (Haken, 1983) to biological phenomena. Hence, it is not at all surprising that advanced methods of mathematics, statistics, and information theory are becoming routine tools in biology (Green et al., 2005), as well as in the related sciences medicine, and agricultural science (Yuan et al., 2008). This is paralleled by a change of view in biology, which is characterized by growing popularity of the notions of network and ecosystem, even beyond neuroscience and microbiology. Not only do biologists nowadays talk about systems biology (Ideker et al., 2001), but also systems (bio)medicine (Lenoir, 1999)(Liu, 2010)(Ayers, 2015)(Maurya, 2010) has become an emerging concept in medicine, pharmacology, and diagnostics (Abu-Asab, 2011).

2 An owerview of quantitative biology

At the beginning of the 20th century, quantitative biology was applied to only a few particular problems, mainly in two different areas, pharmacology on the one hand, and breeding of plants and animals on the other hand.

2.1 Enzyme kinetics

As for pharmacology, probably the first mathematical model of quantitative biology was the Michaelis-Menten theory of enzyme kinetics (Michaelis and Menten, 1913)(Cornish-Bowden, 2013)(Cornish-Bowden, 2015). Decomposing an enzyme's E reaction with a substrate A into the steps of substrate binding, reaction catalysis, and product B release (Schnell, 2014)

$$E + A \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} E{:}A \overset{k_2}{\rightarrow} E + B \tag{1}$$

which can be described by the set of differential equations for E, A, and their complex E:A

$$\frac{d[E]}{dt} = -k_1[E][A] + k_{-1}[E{:}A] + k_2[E{:}A] \tag{2}$$

$$\frac{d[A]}{dt} = -k_1[E][A] + k_{-1}[E{:}A] \tag{3}$$

$$\frac{d[E{:}A]}{dt} = k_1[E][A] - (k_{-1} + k_2)[E{:}A] \tag{4}$$

The corresponding equation for the reaction rates (Michaelis-Menten equation) reads (Pinto and Martins, 2016)

$$v_{initial} = \left(\frac{d[A]}{dt}\right)_{initial} = \frac{A_0 \cdot v_{max}}{K_m + A_0} \tag{5}$$

with A_0 the initial substrate concentration, $v_{max} = k_2 \cdot E_0$, E_0 the initial enzyme concentration, and $K_m = (k_{-1}+k_2)/k_1$, the Michaelis constant. The theory developed by Michaelis and Menten, provided the foundation for quantifying physiology and pharmacology (Dost, 1953). It plays an important role in drug development, still today.

2.2 Quantitative genetics – the breeders' equation

Following the ideas of Darwin and Mendel, people tried to understand and started to predict, thereby optimizing, breeding of plants and animals on the basis of the so-called breeder's equation (Lush, 1937)(Ollivier, 2008). This simple equation allows to estimate the change (response ΔZ) in occurrence of a particular quantifiable trait in the

$$\Delta Z = h^2 \cdot S \tag{6}$$

off-spring generation due to selection for this trait in the parent generation. h^2 is called the heritability, S is the so-called selection differential, which represents the difference of a trait's average value in the respective whole population and the average value in the selected subpopulation. Remarkably, this equation is independent of molecular genomic details, which is the basis for its present-day role in theoretical genomics (Visscher et al., 2008). It should be noted that this early mathematical model being used in quantitative biology, had commercial applications from its very beginning.

2.3 Bioinformatics

The impressive development of DNA, RNA, and peptide sequencing that we see today, was possible only through the collaboration of three disciplines, bioinformatics, computer technology, and sequencing technologies. Bioinformatics provided the mathematical and statistical tools to structure, analyze, and annotate biologically, what had been produced with sequencing machines. High-performance computers were needed, to handle and process the related data, which still today is the core of bioinformatics.

2.4 Molecular phylogeny

Many authors see the actual initialization of bioinformatics in a paper about molecular evolution by Emile Zuckerkandl and Linus Pauling (Zuckerkandl and Pauling, 1962)(Zuckerkandl and Pauling, 1962b)(Zuckerkandl and Pauling, 1965). They recognized the relationship between sequence variation and evolution, defining the foundations of phylogeny, a methodology, still popular today (Lemoine et al., 2018) in sophisticated versions as probabilistic models of evolution.

2.5 Sequence analysis: comparison, alignment, and pattern recognition

The key aspect of sequence analysis is comparison (Pearson and Lipman, 1988) to find and quantify similarity in sequences and accordingly in function. In the beginning, this was based on alignment, the most important methods being the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970) for global alignment and the Smith-Waterman algorithm (Smith and Waterman, 1981) for local alignment. The complications arising from sequence insertions, deletions, and mutations can be managed by statistical scoring systems, which bear some analogy to the concept of entropy in information theory (Altschul, 1991). Thus, scoring not only takes into account identities, but also homologies, i.e. sequence elements, which could be exchanged "easily" in the course of evolution without loss of function, and therefore can be classified as being equivalent. With the growing size of sequence databases, more efficient algorithms, which computationally are less demanding, have been formulated. The most popular one today is the BLAST algorithm (Altschul et al., 1990), providing a quantitative measure for sequence homology in terms of the so-called expectation value (E-value), which is the probability of the respective alignment being purely by chance. The lower the E-value, the more significant is the homology.

Based on pairwise alignment, algorithms for multiple sequence alignment have been developed. From the comparison of a set of DNA or peptide sequences they can generate what is called a consensus sequence, i.e. stretches of sequence that are identical or closely related, while other ranges of the sequences differ more or less significantly by mutations, insertions or deletions. The most widely used software systems today are the different versions of CLUSTAL (W and X) (Larkin et al., 2007).

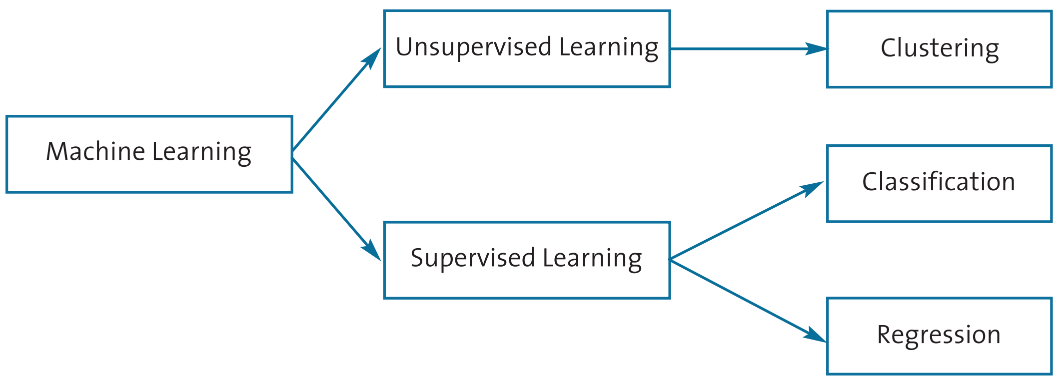
2.6 Pattern recognition in sequence analysis

An important result of sequence comparison is the identification of sequence pattern, which typically provide two kinds of information. Firstly, common, i.e. evolutionary conserved patterns indicate phylogenetic relationships (Fitch and Margoliash, 1967) in a quantitative manner. Secondly, sequence pattern can be correlated with genetic and molecular function. To this end, starting from classification and clustering, algorithms for pattern recognition have been developed (de Ridder et al., 2013). These algorithms also paved the way into machine learning (Baldi and Brunak, 2001) and big data, which was achieved by introducing probabilistic frameworks for the respective models. Building statistical models on the basis of existing data bears a lot of uncertainty, which makes the difference between inference and deterministic conclusion. In statistics, there two different philosophies considered, the Fisher philosophy, also called the frequency approach, and the Bayes philosophy, using prior and posterior distribution knowledge (Leonard and Hsu, 1999), in other words conditional probabilities. It has been shown that Bayesian methods are the most appropriate approach for modeling of biological systems, because it readily allows analyzing data against the background of their actual biological context (Gupta, 2012).

2.7 Some Remarks on Machine Learning

Machine learning is an important aspect of artificial intelligence (Carbonell et al., 1983). The origin of machine learning goes back to the late 50s. It was characterized as a " ... field of study that gives computers the ability to learn without being explic-

Figure 1 Schematic of machine learning. Machine learning can be realized by two different strategies. Unsupervised learning only uses input data and identifies structures in the data. Supervised learning uses training data to create a predictive data model, which subsequently is applied to new input data. Typically, supervised learning is done in a recursive manner, thereby refining the predictive data model more and more (source: Mathworks Inc. 2017).



itly programed" (Samuel, 1959). It is seen as the right choice for solving problems that cannot be tackled by pure computing. This, of course, is not only of interest for the life sciences. The financial industry is using machine learning (Mark et al., 2018) for management of the risks in stock market trading and credit issuing (Fagella, 2018). Furthermore, it is also used in marketing (Chow, 2017), for numerous popular web services (Luckow et al., 2017), in manufacturing, energy production, as well as for automotive, public transport, and aviation maintenance prediction (MathWorks, 2018).

The most important method in what is called unsupervised machine learning is clustering (Filippone et al., 2008), i.e. distinguishing and grouping elements into subsets of a given set of input data based on a measure of similarity applied to the characteristic features of the elements (see fig. 1). With respect to post-processing, this is a divide-and-conquer strategy, because the overall size of a data analysis challenge can be split into analyzing a number of subsets.

Supervised learning is based on experience, i.e. data analyzed for training a particular statistical model obtained through clustering, classification and regression. New input data can then be processed by the trained, predictive model to generate new information. This is done typically in a recursive manner to optimize the parameters of the predictive model.

In general, machine learning is used for data-driven applications and hence requires enormous computing power. For many successful applications, two technological infrastructure innovations have been very important, cloud computing and the involvement of graphical processor units in so-called GPU computing.

2.8 Hidden Markov models

An important machine learning methodology in quantitative biology is realized by so-called Hidden Markov Models (HMMs) (Baldi and Brunak, 2001b). They are tools to analyze serial data like, e.g. time series or biological sequences. In the comparison of sequences (Krogh, 1994), they are used to find relationships between sequences by a probabilistic random walk through a series of states in sequence space (Markov chain). Depending on the selected parameters, new states are either accepted or rejected. Starting from an initial sequence, intermediate sequence states are generated by transitions generated by local repetition, mutation, insertion, and deletion of sequence elements (Fig. 2). Sequence elements can be nucleobases and amino acids, but also sequence pattern like, e.g. base triplets or higher multiplets, amino acid pattern characteristic for a particular folding or function. Accordingly, there are specific transition and emission parameters for sequence elements. In terms of Bayesian probabilities, this gives a quantitative measure of relatedness with respect to the section of the sequence space, reached by the Markov chain.

Usually, a set of refence sequences is taken to train the model and generate its parameters (Rasmussen and Krink, 2003). After optimization the set of parameters obtained in the individual main

Figure 2 Standard Architecture of a HMM for sequence analysis. Each box represents a particular sequence state, derived from the start sequence and generated along a series of transformations, represented by the arrows. The number of boxes in the middle row (backbone) corresponds to the average length of the starting sequences considered. The horizontal arrows in the backbone of the HMM represent a linear Markov chain, a random walk through a series of sequences (the main states), which all have the same length L and differ from their precursor by just one sequence element. Boxes in the upper row correspond to states with deletions of sequence elements, while those in the lower row correspond to insertions of sequence elements. The reflexive circles at the insertion boxes allow variable lengths of insertions through repetition. Each arrow in the schema is associate with a transition probability (source: Baldi and Brunak, 2001).

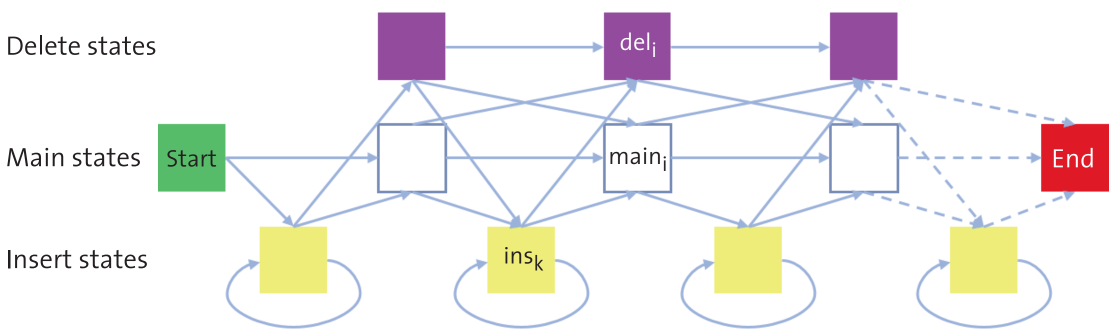
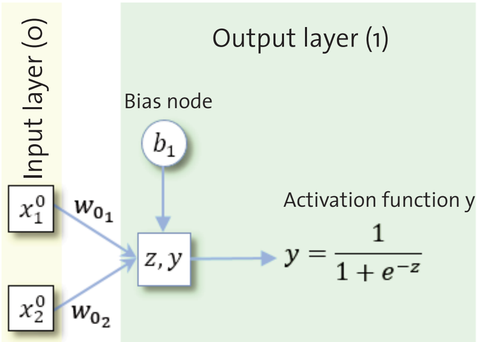


Figure 3 Perceptron. The perceptron is the building block of feed-forward Artificial Neural Networks. Each input value x_i^0 from the nodes in the input layer (0) is multiplied with the specific weight factor w_{0i} for the respective connection to a node in the next layer (1), where it is combined with weighted data from other input nodes to calculate the so-called pre-activation function z. To control the forward feed of the perceptron, a bias node b is added to layer (1). A convolution of z with an appropriate activation function z (here it is the logistic function) produces the output of the shown perceptron (source: own representation).



Pre-activation function z: weighted input +bias

$$z_1 = x_1^0 \cdot w_{0_1} + x_2^0 \cdot w_{0_2} + b_1 = \vec{w}_0 \cdot \vec{x}_1 + b_1$$

and intermediate states carry important information about the sequences involved. Typical applications are the identification of coding areas and protein binding sites of DNA strands.
It should be noted that Hidden Markov Models are also used in speech recognition, optical character recognition, and industrial process control (Windmann et al., 2016). Furthermore, due to their layered structure, Hidden Markov Models are closely related to, or may even be seen as a special case of so-called neural networks, actually one of the most important concepts in machine learning.

2.9 Artificial neural networks, convolutional neural networks, and self-organizing maps

As the name already indicates, artificial neural networks (ANNs) have their origin in the attempt to simulate and understand the characteristics of biological neurons. The human brain is assumed to consist of about 100 x 10^{12} neurons (Herculano-Houzel, 2009) and about ten times as many glial cells, involved in a large number of specific networks by synaptic connections. On arrival at the synapse which is formed by an axon terminal of the emitting neuron and a dendrite of the receiving neuron, and often coupled to a glial cell, the electrical signal is transferred by neurotransmitters and, may be maintained, enhanced, attenuated, or averaged over several signals in the postsynaptic neuron. This behavior has been modeled by so-called perceptrons (McCulloch and Pitts, 1943)(Rosenblatt, 1958)(Stansbury, 2014)(Stansbury, 2014b) (Fig. 3).

Each input value $x_j^{\lambda-1}$ in the input layer (λ-1=0) is multiplied with a weight factor w_j^λ specific for the connection to a node in the next layer λ. In each node, the incoming weighted data items are summed up and can be modified by the parameter b_λ of a so-called bias node. The expression

$$z = \vec{w_\lambda} \cdot \vec{x}_\lambda + b_\lambda \tag{7}$$

is sometimes called the pre-activation function of the signal and is the summation over all data items coming in from all the nodes of the previous layer. The actual activation function is typically a convolution to normalize z to the interval [0,1], which is achieved, e.g. by the sigmoid logistic function

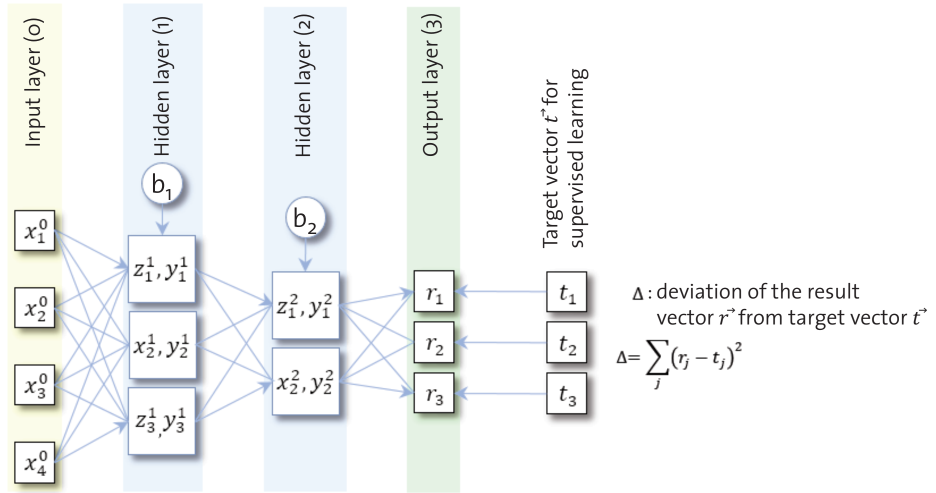
$$y = \frac{1}{1 - e^{-z}} \tag{8}$$

The value of y_j^λ calculated in each node of the layer becomes an input signal for all the nodes of the next layer ($y_j^\lambda \rightarrow x_j^{\lambda+1}$) (Fig. 4).

According to the nature of the problem and the kind of data available, the number of nodes per layer can vary. Each node in layer λ-1 is connected to each node in the next layer λ, and every data item of a given layer is send to all nodes in the successive layer and multiplied with a weight factor, specific for the particular connection. Accordingly, the number of parameters in an ANN is of the order of N·L, with N the average number of nodes per layer and L the number of layers.

Parametrization of ANNs is usually subject to supervised learning. An input vector $\vec{x}$ is taken together with an initial guess of the parameters

Figure 4 An Artificial Neural Network. An artificial neural network in the feed-forward architecture is shown with two hidden layers with 3 and 2 perceptrons, respectively, and one node in the output layer. Training the network by back-propagation requires to minimize the deviation of the result vector $\vec{r}$ from a target vector $\vec{t}$ by gradient-descent of the parameters $\vec{w}_\lambda$ and b_λ for each layer (source: own representation).



$\{\pi=(\vec{w}_\lambda, b_\lambda), 1 \le \lambda \le L\}$ and $\{b_\lambda, 1 \le \lambda \le L\}$ and the output vector of the final layer $\vec{r}$ is calculated. The output is compared to a vector of target values $\vec{t}$. The difference

$$\Delta = \sum_j^N (r_j - t_j)^2 \tag{9}$$

has to be minimized, which can be achieved by the so-called gradient descent method with backpropagation. This requires to calculate the gradients

$$\frac{\partial \Delta}{\partial w_j^\lambda} = 2(r_j - t_j)\frac{\partial r_j}{\partial w_j^\lambda} = 2(r_j - t_j)\frac{\partial y_\lambda}{\partial w_j^\lambda}; j = 1, \cdots N \tag{10}$$

and

$$\frac{\partial \Delta}{\partial b_\lambda} = 2(r_j - t_j)\frac{\partial r_j}{\partial b_\lambda} = 2(r_j - t_j)\frac{\partial y_\lambda}{\partial b_\lambda}; j = 1, \cdots N \tag{11}$$

from which corrections for the respective layer can be determined. Back propagation means that, beginning with the output layer, this procedure has to be repeated for each anterior hidden layer. It should be noted, however, that the gradient descent method is subject to the multiple minima problem. Together with the substantial number of necessary parameters, this is rendering ANNs computationally demanding. Even though, ANNs have a really wide spectrum of applications in science (Musib et al., 2017), medical diagnosis (Kononenko, 2001)(Shen et al., 2017)(Ting et al., 2018), and the industrial context (Lennox et al., 2001).

Recent developments in the field of ANNs go beyond the feed forward architecture. To allow for more flexibility, the number of hidden layers has been augmented. In addition, so-called convolutional neural networks (CNNs) have dropped the restriction of forward transfer and also include connection between nodes within a layer and loops around nodes. The advantage is in the possibility to analyze the data in a hierarchical manner, which is very helpful in image processing and text analysis. Furthermore, CNNs can be used in an unsupervised learning mode (Radford et al., 2016). Working with multi-layered ANNs and CNNs is often called deep learning and has become an integral part of the software infrastructure of many web portals (Hern, 2016)(Abdulkader et al., 2016). It is also the basis of what is called predictive analytics (Siegel, 2016), a methodology that is likely to gain enormous influence on web-based business models. Even though, however, impressive progress in handling complexity has been made, the level reached today is still negligible compared to the complexity of the human brain (Koch, 2012).

Another type of neural networks is the self-organizing map (SOM), which has been inspired by the relationship between an image on the eyes' retina and the corresponding areas in the visual cortex of the brain. Accordingly, SOMs seek to map a dense or contiguous high-dimensional input space to a discrete low-dimensional output space (Kohonen, 1958), thereby compressing information. SOMs belong into the class of non-supervised learning neural networks. In practice one takes a set of nodes representing the input data, the input layer, and maps it to another set of nodes, the computational or output layer. Consequently, the assignment of input nodes to computational nodes is based on competition and collaboration between the computational nodes. In an iterative procedure, the weights and interaction parameters of the individual computational nodes are adjusted to enhance vicinity to the input nodes based on a (projected) distance criterion. Typical applications are the optimization of trajectories for robots (Stergiopoulos, 2012), language recognition, signature recognition, face recognition, seismic data analysis, engineering (Simula et al., 1999), and industrial process control (Frey, 2012). In addition, SOMs have also been used in computer-assisted drug design (Reker et al., 2014) for drug target profiling.

2.10 Computer-assisted molecular design and biomolecular structure prediction

The use of computers in visualizing three-dimensional molecular structures dates back into the 1980s years (Frühbeis et al., 1987). Quantitative structure-activity relationships based on the comparison of molecular structures and their physical, chemical, biological, pharmacological and toxicological properties have been used to design and develop new chemical entities for the respective purposes, ever since (Schneider and Fechner, 2005). Later on, this has been complemented by methods of computer-assisted synthesis planning (Hoffmann, 2009). In parallel, sophisticated algorithms have been developed, which are able to predict the structure of biopolymers. A particular challenge is the assessment of folding and self-organization of the molecules. In principle, so-called ab-initio prediction of structures is possible, but computationally very demanding. The requirements of accuracy of the necessary parameters are enormous. Other approaches start from the prediction of secondary structure elements, whose self-organization is then searched to complete the structures. Quite successful are pragmatic methods, which predict structures on the basis of sequence homology to biopolymers with known tree-dimensional structures (Krieger et al., 2003). The large amount of structures (140591) (RCSB PDB, 2018) published in the RSCB Protein Data Bank (Berman et al., 2018) obtained by x-ray crystallography, multi-dimensional NMR measurements, neutron scattering, and cryo-electron microscopy support this approach significantly. Due to its convenience, homology modeling is widely used in the research and development departments the pharmaceutical industry for what is called structure-based or rational drug design. Drug target structures are used for so-called in-silico screening, which is a computational method of estimating target affinities and rate drug candidates, before they have been synthesized. It can be seen as an option to reduce the amount of chemical syntheses necessary for the development of new drugs (Caldwell, 2015).

2.11 Modeling of biological systems

An important branch of quantitative biology is the modeling of biological systems (Gunawardena, 2014). The targets of modeling range from molecular aggregates to pathways, to cells, to organisms, and populations. The phenomena considered comprise, e.g. material and heat flux balance, metabolic flux analysis, and population dynamics (Shimizu and Matsuoka, 2015). The challenge, but also the motivation, is in modeling and thereby improving comprehension of biological systems' inherent complexity, a situation, also envisioned in medicine (Harz, 2017).

There are basically four different directions in biological systems' modeling. These are (i) the mechanistic modeling of processes at various levels, (ii) the deterministic simulations, using methodologies originating from many-particle physics and fluid-dynamics, (iii) artificial neural and other networks, and (iv) models based on virtual reality, which, by means of man-machine interfaces are supporting human activities and interventions to biological systems.

2.12 Mechanistic modeling and kinetic biological models

The dynamics of a biological system has two aspects, internal dynamics and the interaction and exchange with the system's environment. On a cellular level, this comprises signaling, metabolism, and material transportation. In addition, there is the dynamics of growth, regeneration, and replication, at cellular and organismal level (Chara et al., 2014). Models used in this context are usually systems of ordinary differential equations (ODEs). Examples are kinetic models, and reaction-diffusion models (Britton, 1986)(Volpert and Petrovskii, 2009), also named Turing models (Turing, 1952), which are the most important models to represent the dynamical behavior of living systems (Raue et al., 2013). The dynamics of pattern formation (Kondo and Miura, 2010) and wave propagation, e.g. in signal transduction, gene expression (Gaffney and Monk, 2006), tumor growth, and population growth, can be modeled on the basis of equations like shown in Fig. 5.

A versatile open source software system for

Figure 5 The Reaction-Diffusion Model. The local concentration of a material u is influenced by its formation, degradation, and diffusion. In a multi-component system with $F=F(u,v,w,\cdots)$, the respective differential equations may no longer be separable (source: own representation).

$$\underbrace{\partial u/\partial t}_{\text{Local change of concentration}} = \overbrace{F(u)}^{\text{Formation}} - \underbrace{k \cdot u}_{\text{Degradation}} + \overbrace{D \cdot \Delta u}^{\text{Diffusion}}$$

building and analyzing such kind of models is MORPHEUS (Starruß et al., 2014).

2.13 Deterministic and stochastic modeling

In combination with simulations of the time evolution, deterministic models, which have been developed for molecular or particle dynamics (Vlachakis et al., 2014), and fluid dynamics are used to study, e.g. the system's response to perturbations (Marshall, 2017). For mechanical systems, deterministic models are based on equations of motion that describe the dynamics of the system modeled. By simulation, one obtains a trajectory, documenting the time evolution of the model under the given conditions. Stochastic methods are used for systems with significant noise, which can be represented by random fluctuations. Typical examples are populations (Sharkey, 2011), whose size fluctuates due to death and reproduction.

Rather than looking at a definite trajectory, the Monte-Carlo method (Frenkel, 1990) can be used for scanning the models' phase space by an appropriate random walk, to guarantee ergodicity of the scan. Monte-Carlo methods are often used for high-dimensional systems with many internal interactions.

2.14 Probabilistic biological models

Data-driven biological models are used at all biological levels. For problems that are not clearly deterministic, like establishing relationships between molecular interactions, genetic predisposition, and physiology, models based on Bayesian statistics like HMMs are preferred also in medicine (Couzin, 2004). There are numerous initiatives to take what is called computational medicine to the clinic (Winslow et al., 2012). Models at population level are of particular interest in epidemiology and public health surveillance (Zhang et al., 2013).

2.15 Modeling biological networks

Inside living cells, there are two kinds networks, the network of molecular interactions and the genetic network of genes. While the network of molecular interaction is the physical backbone of cellular functions, the genetic network is an information network (Sharan and Ideker, 2006). Like the genetic code itself, both networks underlie evolutionary changes, and homology of networks is valuable information, e.g. in the field of synthetic biology and in the design of molecular machines. On the basis of graph theory (Friedman, 2004), these networks serve to visualize and structure data of gene expression (Liu, 2018), proteomics, and metabolism. They are the basis for information resources and simulation models for signaling as well as metabolism of cells. Examples are the E_CELL system for generic cell simulations (Tomita et al., 1999), the EcoCyc system (Keseler et al., 2009) for Escherichia coli, and the BioCyc system (Karp et al., 2017) comprising Bacillus subtilis, Saccharomyces cerevisiae, and Homo sapiens (Romero et al., 2004).

2.16 Biological models for virtual and augmented reality

Virtual reality is based on image data in combination of hardware for graphical display, audio, and hardware for haptic interaction and control. Together, this is an example of a man-machine interface and has its roots in flight simulators. But the use of virtual reality has a long tradition also in medicine (Kaltenborn, 1993). It is nowadays well established for surgical education and training (van der Meijden and Schijven, 2009), and also applied in the recovery of stroke patients (Laver et al., 2012)(Henderson et al., 2007) and general traumatic brain injury (Zanier et al., 2018). The underlying data are partly based on geometrical models and partly on photographic images, which, after appropriate image processing are merged with the mathematical model.

Augmented reality is the combination of real-time visual perception with data and information from other sources. A typical example is the head-up display in aircrafts and cars. Information, which usually is not visible while looking out of the front window is projected on the window overlaying the view through the window. This principle is used, e.g. in liver surgery. The problem of liver surgery is related the complex vascular networks of this organ, which can easily be destroyed by surgical interventions, e.g. to remove a tumor, or in liver transplantation. Software systems have been developed that are able to generate a geometrical model of a patient's liver from magnetic resonance tomography (MRT), x-ray computed tomography (CT), positron emission tomography (PET), or ultrasound tomography. The visualized geometrical model of the liver can be used to plan the surgical intervention (Reitinger et al., 2006), and in real-time to support surgeons by projecting the blood vessels onto the surface of the organ (Christ et al., 2017).

3 The present situation

Even though, in the course of the last two centuries, scientists have accumulated a plethora of biological observations, data, and knowledge, numerous open questions and uncertainties are still left (Levin, 2006)(Adams, 2013). Hence, to advance life sciences and its applications, and exploitations, many experts see the necessity of interdisciplinary collaborations of biologists, statisticians, and mathematicians (Hastings, 2005)(Heffelfinger et al., 2004), which actually is the core of data-driven quantitative biology. It is important now, to enable and support such collaborations, not only for academic research, but also for life sciences industries' research and development. It should be noted, however, that such collaborations are not trivial, due to differences in terminology and methodology (Ledford, 2015). In other words, progress based on data and innovative methods, is not an automatism. The way has to be paved (Bialek and Botstein, 2004).

4 Future options for medicine and the healthcare industry

The pharmaceutical industry is in a difficult situation. Therapeutic medical interventions, including medications, can be curative, or can serve for disease maintenance. But they can also be just palliative, if the stability of a patient's status cannot be maintained any longer. In recent decades, this has been a comfortable situation for the pharmaceutical industry in that drugs were administered for an ever-increasing time span between initial diagnosis of an indication and the death of the patient.

The growing financial burden emerging from healthcare systems all over the world (Cunningham, 2010)(Dickman et al., 2016), however, has provoked criticism of the pharmaceutical industry's business model (PriceWaterhouseCoopers, 2009)(Tyson, 2015). In addition, reimbursement of new drugs that the pharma industry is bringing to the market, is more and more coupled to proven superiority with respect to drugs already on the market. Furthermore, agreements on outcome-based reimbursement between pharmaceutical companies and health insurances are becoming more and more common. This puts the pharmaceutical industry under enormous innovation pressure (Taylor, 2015)(Thakor et al., 2017), in particular against the background of difficulties in feeding the R&D pipelines and attrition rates of 80-90% on the way to approval (Caldwell, 2015)(Elsevier white paper, 2017). On the long term it is clear that the pharmaceutical and healthcare industry cannot just continue to optimize existing methodologies and processes. Instead, the data-driven approaches of quantitative biology are an option, to make use of recent scientific progress, changing business models at the same time.

5 Recent scientific progress: options for the healthcare industry

Biological research of recent decades has brought out a number of remarkable achievements. It can be seen already today in science that future progress is coupled to data-driven methodologies of quantitative biology, described in the sections above. Healthcare systems and healthcare industry will have to integrate this kind of methodologies to be able to capture the full potential of the new achievements. Some important examples are given in the following.

Non-coding RNAs that are transcribed from the respective stretches of the genomic DNA, but not further translated into polypeptides exist in virtually all types of cells in all three domains (archaea, prokaryotes, eukaryotes) of biology. A special class of those endogenic RNAs, the micro-RNAs, have been shown to be involved in the regulation of gene expression (Morris, 2008). Being also part of an additional system of inter-cellular, inter-tissue, and inter-species communication, they are released, together with proteins, other types of RNA, and also DNA, in extracellular vesicles, which in turn can be internalized by other cells. The molecular load of such vesicles has been recognized as a source of useful biomarkers and diagnostics for many dysfunctional phenomena and disease states (Mack, 2007)(Wang, et al., 2016).

A special class of small non-coding RNAs are the small inductive RNAs (siRNAs) (Zamore et al., 2000)(Ivanova et al., 2006). In contrast to the microRNAs they are double-stranded and exogenic. In the hands of molecular biologists, they serve as an important possibility for handling and controlling living cells, e.g. stem cells. They too, are a means of controlling gene expression, and probably will belong to new therapeutic tool boxes for future therapeutic concepts in molecular and systems medicine (Wittrup and Lieberman, 2015).

Another important achievement is the possibility to analyze single cells (Wang and Bodovitz, 2010)(Grün and van Oudenaarden, 2015)(Wang and Navin, 2015), comprising genomics, transcriptomics, proteomics, and metabolomics at single cell resolution. Though the majority of applications is in cancer research and has provided deep insight into the cancer genome and tumor development (Zhang et al., 2016), there are also applications in neurology and microbiome research. Altogether, this gives a new perspective for the meaning of precision or evidence-based medicine (Harz, 2017).

The ability, to induce pluripotent human stem cells from adult human skin cells (Takahashi et al., 2007) has opened up entirely new and ethically acceptable perspectives for regenerative medicine (Kang et al., 2016). This has to be seen also against the background of the status of synthetic biology (Cameron et al., 2014), which is about to become an important factor in the synthesis of drugs (Paddon and Keasling, 2014). In addition, particular applications for the medical sector begin to show promising results in diagnostics (Slomovic et al., 2015), the treatment of infectious diseases by bacteriophages and the treatment of cancer by means of engineered bacteria (Ruder et al., 2011), and the use of engineered bacteria (Zhou, 2016) and blood cells (Alapan et al., 2018) as drug carriers.

The spectacular accomplishment of utilizing the prokaryotic immune system for genome editing by means of the CRISPR-Cas9 system and comparable systems (Garrett et al., 2011)(Gaj et al., 2013)(Lee et al., 2018)(Behler et al., 2018), is a breakthrough for synthetic biology and likely to induce a substantial change of paradigm in the future of healthcare. The possibility of directly curing genetic diseases appears in a new light. Not disregarding safety and ethical issues, one has to note that genome editing is about to bring therapeutic interventions to a new level that is likely to reduce the duration of treatments drastically. This should be kept in mind, when talking about the cost of medical genome editing.

The consequent advancement of using large biomolecules for therapeutic purposes is the employment of patients' own modified cells, which is another application of synthetic biology. In the future, autologous stem or progenitor cells, modified by genome editing will be used to treat cancer, genetic diseases, and retroviral infections. Such therapeutic cells are an example of what is called advanced therapy medical products (ATMP) (Hanna et al., 2016). Of course, several hurdles still have to be surmounted. One of them is given by the situation that programming of autologous cells has to be done in a near-patient setting, which typically does not meet GMP requirements and hence will need special attention and care of regulatory agencies (Maciulaitis et al., 2012).

On the other hand, production and routine application of ATMPs requires personnel with new qualifications, different from current medical and healthcare educational profiles. It can be expected that there will be a new kind of industry, let's call it the "advanced therapeutic industry", which will be manufacturer and service provider at the same time. Due to the complexity of the related liability situation, it is not very likely that the traditional "big pharma" industry will be directly involved in this kind of healthcare business.

It is easy to imagine that the new scientific and technological trends, based on data-driven methods, need special expertise. The data scientist will have a key function in future developments (Marx, 2013), be it in a scientific or a commercial environment. Of course, there will be different "flavors" depending on the origin and type of data. Accordingly, besides the qualification in computer science, statistics, and mathematics, data scientists will need further training in the fields of origin of the data. Universities are required to pave the way defining and configuring the respective curricula.

In summary, modern methods and technologies, which have found their way into the life sciences enable to look at living systems with an unprecedented resolution at atomistic, molecular, meso-, and macroscales. At the same time, gigantic amounts of data are generated and need careful data-driven analysis, to really capture the value of these data and to augment knowledge and understanding. For the future, this will be a sound basis for commercial exploitation of the new methodologies and technologies. Let me conclude with a statement by Nobel laureate Richard Feynman who said: "People who wish to analyze nature without using mathematics must settle for a reduced understanding."

References

Abdulkader A, Lakshmirastan A, Zhang J (2016): *Introducing DeepText: Facebook's text understanding engine*, available at https://code.facebook.com/posts/181565595577955/introducing-deeptext-facebook-s-text-understanding-engine/, accessed 21 May 2018.

Abu-Asab M, Chaouchi M, Alesci S, Galli S, Laassri M, Cheema A, Atouf F, VanMeter J, Amri H (2011): Biomarkers in the Age of Omics: Time for a Systems Biology Approach, *OMICS*, **15**, p. 105-112.

Adams M (2013): Open questions: genomics and how far we haven't come, *BMC Biology*, **11**, 109.

Alapan Y, Yasa O, Schauer O, Giltinan J, Tabak A, Sourjik V, Sitti M (2018): Soft erythrocyte-based bacterial microswimmers for cargo delivery, *Science Robotics*, **3**, eaar4423.

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990): Basic Local Alignment Search Tool, *Journal of Molecular Biology,* **215**, p. 403-410.

Altschul S (1991): Amino acid substitution matrices from an information theoretical perspective, *Journal of Molecular Biology*, **219**, p. 555-565.

Ayers D, Day P (2015): Systems Medicine: The Application of Systems Biology Approaches for Modern Medical Research and Drug Development, *Molecular Biology International 2015*, 698169.

Baldi P, Brunak S (2001): *Bioinformatics – The Machine Learning Approach*, MIT Press, Cambridge Mass. p. 43-46.

Baldi P, Brunak S (2001b): *Bioinformatics – The Machine Learning Approach*, MIT Press, Cambridge Mass., p. 166-223.

Behler J, Sharma K, Reimann V, Wilde A, Urlaub H, Hess W (2018): The host-encoded RNase E endonuclease as the crRNA maturation enzyme in a CRISPR–Cas subtype III-Bv system, *Nature Microbiology*, **3**, p. 367-377.

Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P (2000): The Protein Data Bank, *Nucleic Acids Research*, **28**, p. 235-242.

Bernardi G (2012): Fifty-Year Old and Still Ticking.... An Interview with Emile Zuckerkandl on the 50th Anniversary of the Molecular Clock, *Journal of Molecular Evolution,* **74**, p.233-236.

Bialek W, Botstein D (2004): Introductory Science and Mathematics Education for 21st-Century Biologists, *Science*, **303**, p. 788-790.

Britton N (1986): *Reaction Diffusion Equations and Their Applications to Biology*, Academic Press, Burlington.

Caldwell G (2015): In silico tools used for compound selection during target-based drug discovery and development, *Expert Opinion on Drug Discovery*, **10**, p. 901-923.

Cameron D, Bashor C, Collins J (2014): A brief history of synthetic biology, *Nature Reviews Microbiology,* **12**, p.381-390.

Carbonell J, Michalski R, Mitchell T (1983): Machine Learning: A Historical and Methodological Analysis, *AI Magazine*, **4**, p. 69-79.

Chara O, Tanaka E, Brusch L (2014): Mathematical Modeling of Regenerative Processes, in: Galliot B (ed.) *Current Topics in Developmental Biology 108*, Academic Press, Burlington p. 283-317.

Chow, M (2017): *AI and machine learning get us one step closer to relevance at scale*, available at https://www.thinkwithgoogle.com/marketing-resources/ai-personalized-marketing/, accessed 2018-05-10.

Christ B, Dahmen U, Herrmann K, König M, Reichenbach J, Ricken T, Schleicher J, Schwen J, Vlaic S, Waschinsky N (2017): Computational Modeling in Liver Surgery, *Frontiers in Physiology*, **8**, 906.

Cornish-Bowden A (2013): The origins of enzyme kinetics, *FEBS Letters*, **587**, p. 2725-2730.

Cornish-Bowden A (2015): One hundred years of Michaelis–Menten kinetics, *Perspectives in Science*, **4**, p. 3–9.

Couzin J (2004): The New Math of Clinical Trials, *Science*, **303**, p. 784-786.

Cunningham P (2010): The growing financial burden of health care: national and state trends, 2001-2006., *Health Affairs*, **29**, p. 1037-1044.

de Ridder D, de Ridder J, Reinders M (2013): Pattern recognition in bioinformatics, Briefings in Bioinformatics, 14, p. 633-647.

Dickman S, Woolhandler S, Bor J, McCormick D, H. Bor D, Himmelstein D (2016): Health Spending For Low-, Middle-, And High-Income Americans, 1963–2012, *Health Affairs* **35**, p. 1189-1196.

Dost F (1953): G*rundlagen der Pharmakokinetik,* Georg Thieme Verlag, Stuttgart.

Elsevier white paper (2017): *Drug Attrition in Check: Shifting Information Input to Where it Matters*, R&D Solutions for PHARMA & LIFE SCIENCES, Elsevier, Amsterdam.

Fagella D (2018): *Machine Learning in Finance – Present and Future Applications*, available at https://www.techemergence.com/machine-learning-in-finance/, accessed 2018-05-10.

Filippone M, Camastra F, Masulli F, Rovetta S (2008): A survey of kernel and spectral methods for clustering, *Pattern Recognition*, **41**, p. 176-190.

Fitch W, Margoliash E (1967): Construction of Phylogenetic Trees, *Science*, **155**, p. 279-284.

Frenkel D (1990): Monte Carlo Simulations. In:

Catlow C, Parker S, Allen M (eds.) *Computer Modelling of Fluids Polymers and Solids. NATO ASI Series* (Series C: Mathematical and Physical Sciences), **293**. Springer, Dordrecht.

Frey C (2012): *Monitoring of complex industrial processes based on self-organizing maps and watershed transformations,* Proceedings of the 2012 IEEE International Conference on Industrial Technology, IEEE, Piscataway, p. 1041-1046.

Friedman N (2004): Infering Cellular Networks Using Probabilistic Graphical Models, *Science*, **303**, p. 799-805.

Frühbeis H, Klein R, Wallmeier H (1987): Computer-Assisted Molecular Design (CAMD) - An Overview, *Angewandte Chemie Int. Ed. Engl.*, **26**, p. 403-418.

Gaffney E, Monk N (2006): Gene Expression Time Delays and Turing Pattern Formation Systems, *Bulletin of Mathematical biology*, **68**, p. 99-130.

Gaj T, Gersbach C, Barbas III C (2013): ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering, *Trends in Biotechnology*, **31**, p. 397-405.

Garrett R, Vestergaard G, Shah S (2011): Archaeal CRISPR-based immune systems: exchangeable functional modules, *Trends in Microbiology*, **19**, p. 549-556.

Green J, Hastings A, Arzberger P, Ayala F, Kottingham K, Cuddington K, Davis F, Dunne J, Fortin M, Gerber L, Neubert M, (2005): Complexity in Ecology and Conservation: Mathematical, Statistical, and Computational Challenges, *BioScience*, **55**, p. 501-510.

Grün D, van Oudenaarden A (2015): Design and Analysis of Single-Cell Sequencing Experiments, *Cell*, **163**, p. 799-810.

Gunawardena J (2014): Models in biology: 'accurate descriptions of our pathetic thinking', *BMC Biology*, **12**, 29.

Gupta S (2012): Use of Bayesian statistics in drug development: Advantages and challenges, *International Journal of Applied and Basic Medical Research*, **2**, p. 3-6.

Haken H, (1983): *Synergetics – An Introduction,* 3rd ed., Springer, Berlin.

Hanna E, Rémuzat C, Auquier P, Toumi M (2016): Advanced therapy medicinal products: current and future perspectives, *Journal Health Policy and Market Access*, **4**, 31036.

Harz M (2017): Cancer, Computers and Complexity: Decision Making for the Patient, *European Reviews*, **25**, p. 96-106.

Hastings A, Arzberger P, Bolker B, Ives T, Johnson N, Palmer M (2005): Quantitative Bioscience for the 21st Century , *BioScience*, **55** (6), p. 511–517.

Heffelfinger G, Martino A, Gorin A, Xu Y, Rintoul III M, Geist A, Al-Hashimi H, Davidson G, Faulon J, Frink L, Haaland D, Hart W, Jakobsson E, Lane T, Li M, Locascio P, Olken F, Olman V, Palenik B, Plimpton S, Roe D, Samatova N, Shah M, Shoshoni A, Strauss C, Thomas E, Timlin J, Xu D (2004): Carbon Sequestration in Synechococcus Sp.: From Molecular Machines to Hierarchical Modeling, OMICS: *A Journal of Integrative Biology*, **6**, p. 305-330.

Henderson A, Korner-Bitensky N, Levin M, (2007): Virtual Reality in Stroke Rehabilitation: A Systematic Review of its Effectiveness for Upper Limb Motor Recovery, *Topics in stroke Rehabilitation*, **14**, p. 52-61.

Herculano-Houzel S (2009): The human brain in numbers: a linearly scaled-up primate brain, *Frontiers in Human Neuroscience*, **3**, 31.

Hern, A (2016): '*Partnership on AI' formed by Google, Facebook, Amazon, IBM and Microsoft*, available at https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms, accessed 21 May 2018.

Hoffmann R. (2009): *Computer-Aided Synthesis Planning*, in: Elements of Synthesis Planning, Springer, Berlin, p. 145-148.

Howard J (2014): Quantitative cell biology: the essential role of theory, Molecular Biology of the *Cell*, **25**, p. 3438 - 3440.

Ideker T, Galitski T, Hood L, (2001): A New Approach to Decoding Life: Systems Biology, *Annual Review of Genomics and Human Genetics*, **2**, p. 343-372.

Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka I (2006): Dissecting self-renewal in stem cells with RNA interference, *Nature*, **442**, p. 533-538.

Kaltenborn K, Rienhoff O (1993): Virtual Reality in Medicine, M*ethods of Information in Medicine*, **32**, p. 407-417.

Kang H, Shih Y, Nakasaki M, Kabra H, Varghese S (2016): Small molecule–driven direct conversion of human pluripotent stem cells into functional osteoblasts, *Science Advances*, **2**, e1600691.

Karp P, Billington R, Caspi R, Fulcher C, Latendresse M, Kothari A, Keseler I, Krummenacker M, Midford P, Ong Q, Ong W, Paley S, Subhraveti P (2017): *The BioCyc collection of microbial genomes and metabolic pathways,* Briefings in Bioinformatics, bbx085, https://doi.org/10.1093/bib/bbx085.

Keseler I, Bonavides-Martínez C, Collado-Vides J, Gama-Castro S, Robert P. Gunsalus R, Johnson D, Krummenacker M, Nolan L, Paley S, Paulsen I, Peralta-Gil M, Santos-Zavaleta A, Shearer A, Karp P (2009): EcoCyc: A comprehensive view of Escherichia coli biology, *Nucleic Acid Research*, **37**, p. D464-D470.

Koch C (2012): Modular Biological Complexity, *Science*, **337**, p. 531-532.

Kohonen T (1958): Self-Organized Formation of Topologically Correct Feature Maps, Biological *Cyber-*

netics, **43**, p. 59-69.

Kondo S, Miura T (2010): Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation, *Science*, **329**, p. 1616-1620.

Kononenko I (2001): Machine learning for medical diagnosis: history, state of the art and perspective, *Artificial Intelligence in Medicine*, **23**, p. 89-109.

Krieger E, Nabuurs S, Vriend G (2003): Introduction to homology modeling, *Methods of Biochemical Analysis*, **44**, p. 509-523.

Krogh A, Brown M, Mian I, Sjölander K, Haussler D (1994): Hidden Markov Models in Computational Biology Applications to Protein Modeling, *Journal of Molecular Biology*, **235**, p. 1501-1531.

Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace I, Wilm A, Lopez R, Thompson J, Gibson T, Higgins D (2007): Clustal W and Clustal X version 2.0, *Bioinformatics,* **23**, p. 2947-2948 (2007).

Laver K, George S, Thomas S, Deutsch J, Crotty M (2012): Cochrane review: virtual reality for stroke rehabilitation, European *Journal of Physical and Rehabilitation Medicine*, **48**, p. 523-530.

Ledford H (2015): Team Science, *Nature*, **525**, p. 308-311.

Lee H, Zhou Y, Taylor D, Sashital D (2018): Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays, *Molecular Cell*, **70**, p. 1-12.

Lemoine F, Domelevo Entfellner J, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O (2018): Renewing Felsenstein's phylogenetic bootstrap in the era of big data, *Nature*, **556**, p. 452-456.

Lennox B, Montague G, Frith A, Gent C, Bevan V (2001): Industrial application of neural networks - an investigation, *Journal of Process Control*, **11**, p. 497-507.

Lenoir T (1999): Shaping Biomedicine as an Information Science, in: Bowden M, Bellardo Hahn T, Williams R (eds.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems, ASIS Monograph Series*, Information Today, Inc., Medford, New Jersey, p. 27-45.

Leonard T, Hsu J (1999): *Bayesian Methods – An Analysis for Statisticians and Interdisciplinary Researchers,* Cambridge University Press, Cambridge UK.

Levin S (2006): Fundamental Questions in Biology, *PLoS Biology*, **4**, e300.

Liu E (2010): Foundations for Systems Biomedicine: an Introduction, in: Liu E, Douglas A. Laufenburger D, (eds.), *Systems Biomedicine – Concepts and Perspectives*, Academic Press, Cambridge MA, p. 3-13.

Liu Z (2018): Towards precise reconstruction of gene regulatory networks by data integration, *Quantitative Biology*, https://doi.org/10.1007/s40484-018-0139-4.

Luckow A, Cook M, Ashcraft N, Weill E, Djerekarov E, Vorster B (2017): *Deep Learning in the Automotive Industry: Applications and Tools*, arXiv:1705.00346v1 [cs.LG].

Lush J (1937): *Animal Breeding Plans*, Iowa State College Press, Ames, Iowa.

Maciulaitis R, D'Apote L, Buchanan A, Pioppo L, Schneider C (2012): Clinical Development of Advanced Therapy Medicinal Products in Europe: Evidence That Regulators Must Be Proactive, *Molecular Therapy*, **20**, p. 479-482.

Mack, G (2007): MicroRNA gets down to business, *Nature Biotechnology*, **25**, p. 631-638.

Mark C, Metzner C, Lautscham L, L. Strissel P, Strick R, Fabry B (2018): Bayesian model selection for complex dynamic systems, *Nature Communications*, **9**, 1803.

Marshall W (2017): *Introduction to Quantitative Cell Biology, Wallace F. Marshall*, Editor, Morgan & Claypool, San Francisco.

Maurya M, Subramaniam S (2010): Computational Challenges in: Systems Biology, in: Liu E, Laufenburger D, (eds.), *Systems Biomedicine – Concepts and Perspectives*, Academic Press, London.

Marx V, (2013): Biology: The big challenges of big data, *Nature*, **498**, p. 255-260.

MathWorks, Inc. (2018): H*ow Machine Learning Works*, available at https://www.mathworks.com/discovery/machine-learning.html#how-it-works, accessed 2018-05-10.

Mayer E, (1997): *This is Biology - The Science of the Living World*, Harvard University Press, Cambridge MA.

McCulloch W, Pitts W (1943): A logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics,* **5**, p. 115-133.

Michaelis L, Menten M (1913): Kinetik der Invertinwirkung, *Biochemische Zeitung*, **49**, p. 333-369.

Morris, K (2008): in: Morris, K (ed.), *RNA and the Regulation of Gene Expression - A Hidden Layer of Complexity*, Caister Academic Press, Poole.

Musib M, Wang F, Tarselli M, Yoho R, Yu K, Andrés R, Greenwald N, Pan X, Lee C, Zhang J, Dutton-Regester K, Johnston J, Sharafeldin I (2017): Artificial intelligence in research, *Science*, **357**, p. 28-30.

Needleman S, Wunsch C (1970): A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins, *Journal of Molecular Biology,* **48**, p. 443-453.

Ollivier L (2008): Jay Lush: Reflections on the past, *Lohmann Information*, **43**, p. 3-12.

Paddon C, Keasling J (2014): Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development, *Nature Reviews Microbiology*, **12**, p. 355-367.

Pearson W, Lipman D (1988): *Improved tools for* *biological sequence comparison*, Proc. Natl. Acad. Sci. USA, 85, p. 2444-2448.

Pinto M, Martins P (2016): In search of lost time constants and of non-Michaelis—Menten parameters, *Perspectives in Science*, **9**, p. 8-16.

PricewaterhouseCoopers (2009) *Pharma 2020: Challenging business models*, PWC International Ltd.

Radford A, Metz L, Chintala S (2016)*: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*, arXiv:1511.06434v2, 7 Jan 2016.

Rasmussen T, Krink T (2003): Improved Hidden Markov Model training for multiple sequence alignment by a particle swarm optimization—evolutionary algorithm hybrid, *BioSystems*, **72**, p. 5-17.

Raue A, Schilling M, Bachmann J, Matteson A, Schelke M, Kaschek D, Hug S, Kreutz C, Harms B, Theis F, Klingmüller U, Timmer J (2013): Lessons Learned from Quantitative Dynamical Modeling in *Systems Biology*, PLoS ONE, **8**, e74335.

RCSB PDB (2018): *Research Collaboratory for Structural Bioinformatics*, available at https://www.rcsb.org, accessed 2018-05-12.

Reitinger B, Bornik A, Beichel R, Schmalstieg D (2006): Liver Surgery Planning Using Virtual reality, *IEEE Computer Graphics and Applications*, **26**, p. 36-47.

Reker D, Rodrigues T, Schneider P, Schneider G (2014): Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus, *Proc. Natl. Acad. Sci. USA*, **111**, p. 4067-4072.

Robertson M (2013): Open questions in biology - A tenth anniversary series, *BMC Biology*, **11**, 7.

Romero P, Wagg J, Green M, Kaiser D, Krummenacker M, Karp P (2004): Computational prediction of human metabolic pathways from the complete human genome, *Genome Biology*, **6**, R2.

Rosenblatt F (1958): The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, *Psychological Review*, **65**, p. 386-408.

Ruder W, Lu T, Collins J (2011): Synthetic Biology Moving into the Clinic, *Science*, **333**, p. 1248-1252.

Samuel A. (1959): Some studies in machine learning using the game of checkers. *Recent Progress II, IBM Journal of Research and Development*, **3**, p. 535-554.

Schneider G, Fechner U (2005): Computer-Based De Novo Design of Drug-Like Molecules, *Nature Reviews Drug Discovery*, **4**, p. 649-662.

Shen D, Wu G, Suk H (2017): Deep Learning in Medical Image Analysis, *Annual Review of Biomedical Engineering*, **19**, p. 221-248.

Schnell S, (2014): Validity of the Michaelis–Menten equation – steady-state or reactant stationary assumption: that is the question, *FEBS Journal*, **281**, p. 464-472.

Schölkopf B, Smola A (2002): *Learning with Kernels*, MIT Press, Cambridge MA.

Sharan R, Ideker T (2006): Modeling cellular machinery through biological network comparison, *Nature Biotechnolgy*, **24**, p. 427-433.

Sharkey K (2011): Deterministic epidemic models on contact networks: Correlations and unbiological terms, *Theoretical Population Biology*, **79**, p. 115-129.

Shimizu K, Matsuoka Y (2015): Fundamentals of Modeling of Biosystems, in: Shimizu K, Matsuoka Y (eds.) *Fundamentals of Systems Analysis and Modeling of Biosystems and Metabolism*, Bentham Science Publishers Ltd., Sharja U.A.E., p. 81-132.

Siegel E, (2106): *Predictive Analytics, Wiley*, Hoboken.

Simula O, Vesanto J, Alhoniemi E, Hollmén J (1999): Analysis and Modeling of Complex Systems Using the Self-Organizing Map, in: Kasabov, Nikola, Kozma, Robert (Eds.), *Neuro-Fuzzy Techniques for Intelligent Information Systems*, Physica-Verlag, Heidelberg, pp. 3-22.

Slomovic S, Pardee K, Collins J (2015): Synthetic biology devices for in vitro and in vivo diagnostics, *Proc. Natl. Acad. Sci. USA*, **112**, p. 14429–14435.

Smith T, Waterman M (1981): Identification of Common Molecular Subsequences, *Journal of Molecular Biology*, **147**, p. 195-197.

Stansbury, D (2014): *A Gentle Introduction to Artificial Neural Networks,* available at https://theclevermachine.wordpress.com/2014/09/11/a-gentle-introduction-to-artificial-neural-networks/, accessed 2018-05-12.

Stansbury, D (2014b): *Error Backpropagation & Gradient Descent for Neural Networks*, available at https://theclevermachine.wordpress. com/2014/09/06/derivation-error-backpropagation-gradient-descent-for-neural-networks/, accessed 2018-05-12.

Starruß J, de Back W, Brusch L, Deutsch A (2014): Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology, *Bioinformatics*, **30**, p. 1331-1332.

Stergiopoulos Y, Kantaros Y, Tzes A (2012): *Connectivity-aware coordination of robotic networks for area coverage optimization*, Proceedings of the 2012 IEEE International Conference on Industrial Technology, IEEE, Piscataway, p. 31-35.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007): Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors, *Cell*, **131**, p. 861-872.

Taylor D (2015): The Pharmaceutical Industry and the Future of Drug Development, in: Hester R, Harrison R (eds.), *Pharmaceuticals in the Environment,* *Royal Chemical Society*, Cambridge UK, p. 1-33.

Thakor R, Anaya N, Zhang Y, Vilanilam C, Siah K, Wong C, Lo A (2017): Just how good an investment is the biopharmaceutical sector?, *Nature Biotechnology*, **35**, p. 1147-1157.

Ting D, Liu Y, Burlina P, Xu X, Bressler N, Wong T (2018): AI for medical imaging goes deep, *Nature Medicine*, **24**, p. 534-540.

Tomita M, Hashimoto K, Takahashi K, Shimizu T, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter C, Hutchison III C (1999): E-CELL: software environment for whole-cell simulation, *Bioinformatics*, **15**, p. 72-84.

Turing A (1952): The chemical basis of morphogenesis, *Philosophical Transactions of the Royal Society London*, B **237,** p. 37–72.

Tyson B (2015): *Why Pharma Must Change Its Model, Forbes Pharma & Healthcare,* July 30, 2015, available at: https://www.forbes.com/sites/matthewherper/2015/07/30/why-pharma-must-change-its-model/2/#7dd1b84e3192, accessed 23 March 2018.

van der Meijden O, Schijven M (2009): The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review, *Surgical Endoscopy*, **23**, p. 1180–1190.

Venter C, Cohen D, (2004): The Century of Biology, *New Perspectives Quarterly*, **21**, p. 73-77.

Visscher P, Hill W, Wray N, (2008): Heritability in the genomics era — concepts and misconceptions, *Nature Reviews Genetics*, **9**, p. 255-266.

Vlachakis D, Bencurova E, Papangelopoulos N, Kossida S (2014): Current State-of-the-Art Molecular Dynamics Methods and Applications, in: Donev R (ed.), *Advances in Protein chemistry and Structural Biology 94*, Academic Press, Cambridge MA, p. 269-313.

Volpert V, Petrovskii S (2009): Reaction-diffusion waves in biology, *Physics of Live Reviews*, **6**, p. 267-310.

Voosen P (2017): The AI Detectives, *Science*, **357**, p. 22-27.

Wang G (2017): Global quantitative biology can illuminate ontological connections between diseases, *Quantitative Biology,* **5**, p. 191–198.

Wang D, Bodovitz S (2010): Single cell analysis: the new frontier in 'omics', *Trends in Biotechnology*, **28**, p. 281-290.

Wang J, Chen J, Sen S (2016): MicroRNA as Biomarkers and Diagnostics, *Journal of Cellular Physiology*, **231,** p. 25-30.

Wang Y, Navin N (2015): Advances and Applications of Single-Cell Sequencing Technologies, *Molecular Cell*, **58**, p. 598-609.

Windmann S, Eickmeyer J, Jungbluth F, Badinger J, Niggemann O (2016): Monitoring of Complex Industrial Processes based on Self-Organizing Maps and Watershed Transformations, in: Niggemann O, Beyerer J (eds.), *Machine Learning for Cyber Physical Systems*, Springer-Verlag, Berlin, p. 45-50.

Winslow R, Trayanova N, Geman D, Miller M (2012): Computational Medicine: Translating Models to Clinical Care, *Science Translational Medicine* **4**, 158rv11.

Witkowski J, (2018): *Cold Spring Harbor Symposium on quantitative Biology 1933: Surface Phenomena*, available at http://symposium.cshlp.org/site/misc/topic1.xhtml, accessed 21 May 2018.

Wittrup A, Lieberman J (2015): Knocking down disease: a progress report on siRNA therapeutics, *Nature Reviews Genetics*, **16,** p. 543-552.

Yuan H, Xiao S, Wang Q, Wu K (2008): A Bioeconomic Model by Quantitative Biology to Estimate Swine Production. In: Li D (ed.) *Computer And Computing Technologies In Agriculture*, Volume I. CCTA 2007. The International Federation for Information Processing, vol 258. Springer, Boston, MA, p. 667-675.

Zamore P, Tuschl T, Sharp P, Bartel D (2000): RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals, *Cell*, **101**, p. 25-33.

Zanier E, Zoerle T, Di Lernia D, Riva G (2018): Virtual Reality for Traumatic Brain Injury, *Frontiers in Neurology*, **9**, 345.

Zhang M, Tang C, (2013): QB: A new inter- and multi-disciplinary forum for modeling, engineering and understanding life, *Quantitative Biology* **1**, p. 1-2.

Zhang X, Marjani S, Hu Z, Weissman S, Pan X, Wu S (2016): Single-Cell Sequencing for Precise Cancer Research: Progress and Prospects, *Cancer Research*, **76**, p. 1305-1312.

Zhou S (2016): Bacteria synchronized for drug delivery, *Nature*, **536**, p. 33-34.

Zuckerkandl E, Pauling L (1962): Molecular disease, evolution, and genetic heterogeneity. In: M. Kasha, B. Pullman (eds.), *Horizons in biochemistry, academic Press*, New York, p. 189-225.

Zuckerkandl E, Pauling L (1962b): Evolutionary Divergence and Convergence in Proteins, in: Bryson V, Henry Vogel H (eds.), *Evolving Genes and Proteins* (New York: Academic Press, 1962), p. 97-166.

Zuckerkandl E, Pauling L (1965): Molecules as Documents of Evolutionary History, *Journal of Theoretical Biology*, **8**, p. 357-366.

Submission guidelines

Manuscripts may be submitted for consideration as research papers, papers for the practitioner's section or as commentaries.

All submitted manuscripts should contain original research not previously published and not under consideration for publication elsewhere. Papers may come from any country but must be written in American English.

Initial Submission

Authors are required to submit manuscripts via e-mail (submit@businesschemistry.org). Please prepare the text in Microsoft Word or rtf-format. When submitting a manuscript, please include the following information:

- Information about the authors (affiliation, postal address, e-mail address)
- Tables and graphics separately in jpg-format (high quality), Microsoft Excel or Powerpoint.

Additionally, please stick to the formal requirements presented below, especially concerning citations and graphics. Manuscripts disregarding the guidelines may be returned for revision prior to any reviewing activity.

Organization of the manuscript

Manuscripts can be arranged in the following order:

- Title, author(s), and complete name(s) of institution(s), corresponding author's e-mail address
- Abstract
- Introduction
- Methods
- Results
- Discussion
- References

These guidelines are, however, flexible, especially for case studies. To structure your manuscript, please try to restrict yourself to a maximum of three levels of headlines.

References and footnotes

The authors are fully responsible for the accuracy of the references. Citations in the text contain only authors' names and date of publication [e.g. (Leker, 2001), (Bröring and Leker, 2006) or (Bröring et al., 2006) when three or more authors]. Full references have to be included at the end of the paper in alphabetical order. For more information on the reference style, please visit www.businesschemistry.org.

Tables and figures

Tables must have titles and sufficient empirical detail in a legend immediately following the title to be understandable without reference to the text. Each column in a table must have a heading, and abbreviations, when necessary, should be defined in the legend. Please number the tables. Figures should have titles and explanatory legends containing sufficient detail to make the figure easily understood. Appropriately sized numbers, letters, and symbols should be used. The abscissa and ordinate should be clearly labeled with appropriately sized type.

Revision

Revise text in Microsoft Word. Revise graphics at publication quality resolution. You may submit the revised manuscript as a single Microsoft Word document. Please send the revised manuscript via e-mail to the Editor who contacted you. You will need:

- Your submission number
- A cover letter with information for the Executive Editor and responses to raised concerns
- The revised manuscript.

Publication

The Executive Editor responsible for your submission might ask you to change the format of your files in order to publish it. If the manuscript does not fulfill the formal requirements, the paper might be denied publication.

Comments

More information can be found on www.businesschemistry.org. If you have any further questions or comments you are welcome to contact us at contact@businesschemistry.org.

Thank you for your contribution!