Blitzlichter vom Kongress „BioData World Congress“ in Basel

Goldgräberstimmung bei den Datenschatzsuchern

Wer in Deutschland an „Big Data“ oder „Künstliche Intelligenz“ denkt, hat vielleicht zuerst Prof. Dr. Gerd Antes „Spurious Correlations“ (MVF 05/18) im Sinn. Oder an den in diesem Zusammenhang gern benutzten und sicher auch nicht falschen Satz „Garbage in, garbage out“, den der IBM-Programmierer George Fuechsel so um das Jahr 2004 in die Welt gesetzt hat. Seitdem hat sich viel getan, wohl mehr als in all den Jahren davor, was insbesondere seit 2007 gilt, als das erste iPhone das Licht der Welt erblickte: Seitdem verdoppelt sich das verfügbare Wissen Jahr für Jahr. Einen tiefen Einblick in das, was im Bereich der Gesundheitsbranche passiert, gibt seit 2004 der zuerst in Cambridge, dann in Basel stattfindende „BioData World Congress“, der sich selbst stolz als „Europas größter Kongress zu Big Data in der pharmazeutischen Entwicklung und im Gesundheitswesen“ bezeichnet. 2021 besuchten ihn immerhin rund 2.000 Teilnehmer, die den Worten von über 350 Vortragenden lauschten, um – so die Kongressankündigung – „part of the story“ zu sein. Frappierend dabei: Teilnehmer wie Redner kamen fast aus der ganzen Welt, nur aus Deutschland waren ganz, ganz wenige dabei. Dabei hätte es sich gelohnt.

>> Das hehre Versprechen der veranstaltenden Agentur, die Londoner Terrapinn Holdings Limited, dass dieser „Weltklasse-Kongress Innovationen präsentieren, Erfolge demonstrieren und Hindernisse und Barrieren überwinden“ werde, um sicherzustellen, dass die Innovationen in der Genomik und „Big Data“ schnell und effizient in die Klinik gelangen, mag Marketingsprech zuzuschreiben sein. Vor-Ort in Basel gab es jedoch ein regelrechtes Stelldichein vor allem der international tätigen Pharmabranche, die auch ein Gros der Vortragenden stellte. Mit oft eigenartigen, hierzulande (auch in deutscher Übersetzung) unüblichen Job-Bezeichnungen wie „Population Analytics“ (Janssen), „External Engagement for Clinical Digital Innovation“ (Sanofi), „Head of Costumer Succcess“ (Novartis), eher bekannteren wie „Tech Head of Innovation and Data Science“ (GSK), oder schon fast eingedeutschten wie „Head of IT“ (GSK) und „Chief Strategy Officer“ (Eleven Therapeutics). Mit dabei waren in Deutschland meist (noch) absolut unbekannte international tätige IT- und Wissenstechnologie-Firmen, wie Sema 4 (1), 54gene (2), Benchling (3), GenomSys (4) oder DNAnexus (5). Bei einigen der Genannten handelt es sich um hochinteressante Innovateure, wie etwa Abzu (6), Systemhäuser aus den verschiedensten Themenbereichen wie Sevenbridges (7), die sich als „scientific discoverer and accelerator“ verstehen, oder die deutsche Firma metaphacts (8), die sich „als end to end-realizer für Graph- und Semantic-Anwendungen“ einen Namen gemacht hat.

Neben Marktgrößen wie BC Platforms (9) und AWS (10) fand sich eine Vielfalt von passgenauen und offensichtlich auch schon längst erprobten Angeboten für Warehousing wie Databricks (11) sowie Ontologie-Entwickler, Textmining- und Graph-Datenbank-Anbieter wie Ontochem (12) oder Genestack (13) ein. Allesamt erwecken sie durchaus den Eindruck einer wohlsortierten, zuverlässigen und erfahrenen Dienstleisterstruktur. Die Berater-Gilde – angeführt durch IQVIA (14) mit alleine zwei Ständen – und Tochterunternehmen internationaler Medizinverlage wie SciBite von Elsevier rundeten mit ihrem Ansatz „content to data“ (15) das Bild eines bereits sehr hoch elaborierten Marktes ab. Eines Markts, der eines im absoluten Mittelpunkt hat: Daten, Daten und nochmals Daten. Hier werden Daten als eigenes Produkt mit eigener Werthaltigkeit verstanden (Abb. 1).

Voraussetzung dafür ist allerdings, dass eine hohe Kompetenz in Data Science nicht nur entwickelt wird, sondern zunehmend als ein eigener Wert verstanden werden muss, der die integrierte Bewirtschaftung des gesamten Datenuniversums eines Unternehmens als zentrales Asset der Unternehmensführung aufbaut, steuert und – das wäre zu hoffen – vor allem zum Wohle des Patienten einsetzt.

Dieses hier sichtbar werdende Unwohlsein hat einen Grund: Denn wohl am frap-

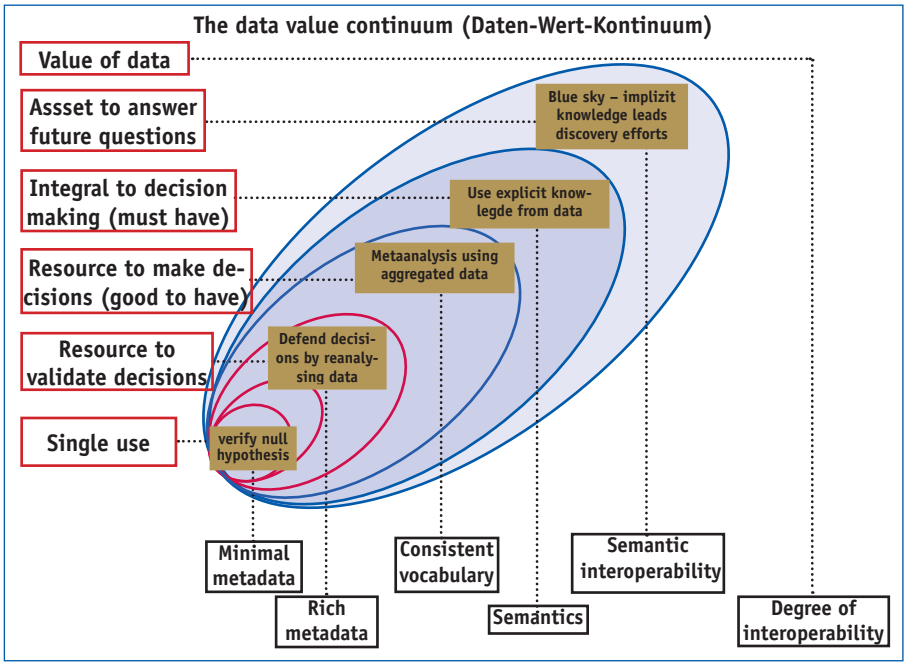


Abb. 1: Daten-Wert-Kontinuum (The data value continuum); Vortrag AstraZeneca, BioData World Congress 2021.

Links

1: https://sema4.com
2: https://54gene.com
3: https://www.benchling.com
4: https://genomsys.com
5: https://www.dnanexus.com
6: https://www.abzu.ai
7: https://www.sevenbridges.com
8: https://metaphacts.com
9: https://www.bcplatforms.com
10: https://aws.amazon.com/de/
11: https://databricks.com
12: https://ontochem.com
13: https://genestack.com
13: https://www.iqvia.com
14: ttps://www.scibite.com

pierensten war die Art und Weise, wie in vielen Vorträgen mit dem in Deutschland so hoch gehandelten und so gut geschützten Gut „Daten" umgegangen wurde. So lauteten einige sehr tief blicken lassende Zitate aus Vorträgen internationaler (wohlgemerkt: nicht deutscher) Industrievertreter, die nur pars pro toto stehen können, weil es schier unmöglich war, in drei Kongresstagen über vielfach parallel laufenden Sessions auch nur annähernd den Überblick (Abb. 1-3) zu behalten: „Maximierung der Datennutzung" (Merck), „Vertrauenswürdige Partnerschaften auf Basis von RWE & Daten sind der Kern der Arbeit" (AstraZeneca) oder „offensiver Einsatz von RWD/RWE auch in Rare Diseases, also auch bei schmaler und inkonsistenter Datenlage" (Ipsen).

In weiteren Vorträgen stellte Sanofi seine „strategische Leitlinie" unter dem Motto „Fairify your own data and Ecosystem" vor, während Novo Nordisk ein „Data Value Continuum – zentralisiertes Ontologie-Managementsystem" beschwor und Bayer einen „DCT Walk" (Abb. 2) vorstellte, wobei „Decentralized Clinical Trials" die zentralen Herausforderung hätten, „verstreute regulatorische Akzeptanz in verschiedenen Ländern" zu erzeugen. So ziemlich alle Industrievertreter proklamieren umfassende Daten und Wissensmodelle, basierend auf Technologien aus den Bereichen der AI/Graph/ML.

Diese werden schon heute durch weltweite Datenstandards unterlegt wie die Grundlagen von F:A:I:R:, wobei diese sogar schon auf die EMA-Frameworks sowie auf CTIS, IDMP, SPOR, Elixir, IMI etc. ausgelegt sind. Ergänzend dazu haben sich starke Allianzen in der Industrie gebildet, wie etwa Pistoia, GA4GH, AAIH* usw.

Das hier und da durchscheinende Szenarium des AI- und datadriven Lifecycle-Managements – international, intersektoral und obendrein interdisziplinär – könnte man durchaus als ein künftig weltumspannendes Modell der Pharma-Industrie zur Beherrschung der Gesundheitsmärkte verstehen. Man muss das natürlich immer im Licht ganz anderer Datenschutzregeln, als jenen die in Deutschland gelten, verstehen und auch einordnen. Dennoch entwickelte sich dieser Fachkongress zu einer regelrechten Wissens-Plattform, die es durchaus erlaubte zu erkennen, wohin sich die Bereiche „Big Data", „Künstliche Intelligenz" und „Wissensmanagement" nicht nur hin entwickeln werden, sondern das längst getan haben. Oder auch, um die großen und kategorischen Veränderungen zu erfassen, die sich durch den mittlerweile konzertierten, digitalen Wandel in der gesamten Gesundheitswirtschaft einstellen werden.

Nur erkennt man dieses Faktum in Deutschland noch viel zu wenig. Sicher auch deshalb, weil deutsche Besucher und Vortragende auf dem Kongress selten anzutreffen waren, wobei Ausnahmen die Regel bestätigten, wie etwa einige Topmanager deutscher Pharmafirmen, Universitäten und Unikliniken oder Forschungsinstitute wie das Fraunhofer zeigten.

Geprägt wurde der Kongress indes durch die Besucher und Experten der internationalen Industrie, der großen US-Universitäten oder europäischen KI-Hotspots wie Cambridge. Sie trafen auf diesem Spielplatz der Ideen und Möglichkeiten auf viele Start-ups, meist geführt von Acceleratoren, sowie auf etablierte Datendienstleister und viele, durchaus bereits erfolgreiche Newcomer, die ein kompaktes und hochkarätiges Programm genießen konnten.

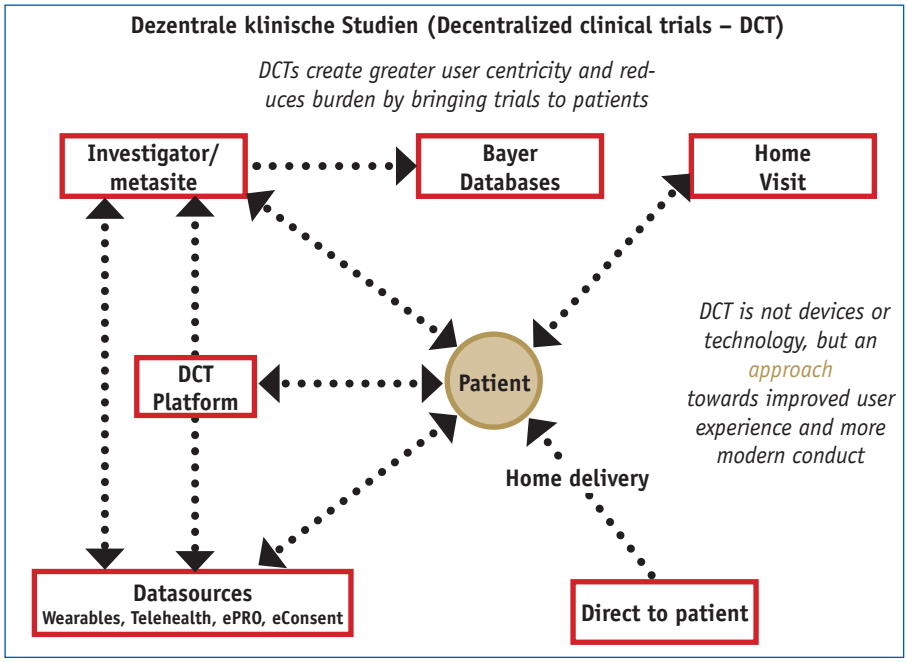


Abb. 2: Dezentrale klinische Studien (Decentralized clinical trials), Vortrag Bayer, BioData World Congress 2021.

Schwieriger Einstieg in die Szene

Allerdings ist ein Einstieg in das Gesamtszenario nicht mehr niederschwellig möglich, weil die Komplexität und der hohe methodische Organisationsgrad diese Wissensuniversen zu Alleinstellungsmerkmalen der großen Marktteilnehmer macht, aber gleichzeitig Barrieren für Sprunginnovationen, Patienteninitiativen oder auch für alternative Gesundheitsmodelle schafft. Getrieben von der Industrie, unterstützt von der digitalen technologischen Elite und obendrein ausgestattet anscheinend ohne jegliche finanzielle Begrenzungen, entsteht hier eine daten- und informationsgetriebene „next generation healthcare". Der damit mögliche Wandel wird in den Dimensionen Systematik, Technologien und Methoden massive Auswirkungen auf jedes Gesundheitssystem eines jeden Staates sowie deren handelnden Personen und Organisationen haben.

Dabei agiert die Industrie schon jetzt – so stellte es sich auf diesem Weltkongress zumindest dar – mit einem umfassenden Steuerungsanspruch: von der Forschung bis zur realen Versorgung. Auch weil es die Industrie – trotz allem Wettbewerb – geschafft hat, miteinander zu kollaborieren, Standards zu setzen oder maßgeblich zu beeinflussen, um über institutionalisierte Kooperationen neue gemeinschaftliche Marktansätze (Bsp.: Pistoia Alliance, GA4GH, AAIH)* zu finden und zu etablieren. Dabei ist zumindest in Europa alles Handeln schon auf die künftigen wissenschaftlichen und regulatorischen Rahmenbedingungen rund um die EMA ausgerichtet, d. h. Standards und Initiativen (wie CTIS, IDMP, SPOR, Elixir, IMI usw.*) sind „internalisiert". Das Ganze fußt auf beeindruckenden und faszinierenden technologischen und methodischen Ansätzen, deren Translation in die Wissenschaft und auch den Alltag unumkehrbar vorbestimmt ist.

Kritisch dabei ist sicher, dass die fast ausschließliche Beteiligung kommerzieller Marktteilnehmer an den Innovationen und der Modellbildung bedenkliche Folgen für die Ausgewogenheit im Bereich der Versorgung auf die Daseinsvorsorge haben könnte. Wenn

Zitationshinweis

Klein, M.: „Goldgräberstimmung bei den Datenschatzsuchern“, in „Monitor Versorgungsforschung“ (06/21), S. 24-26. http://doi.org/10.24945/MVF.06.21.1866-0533.2355

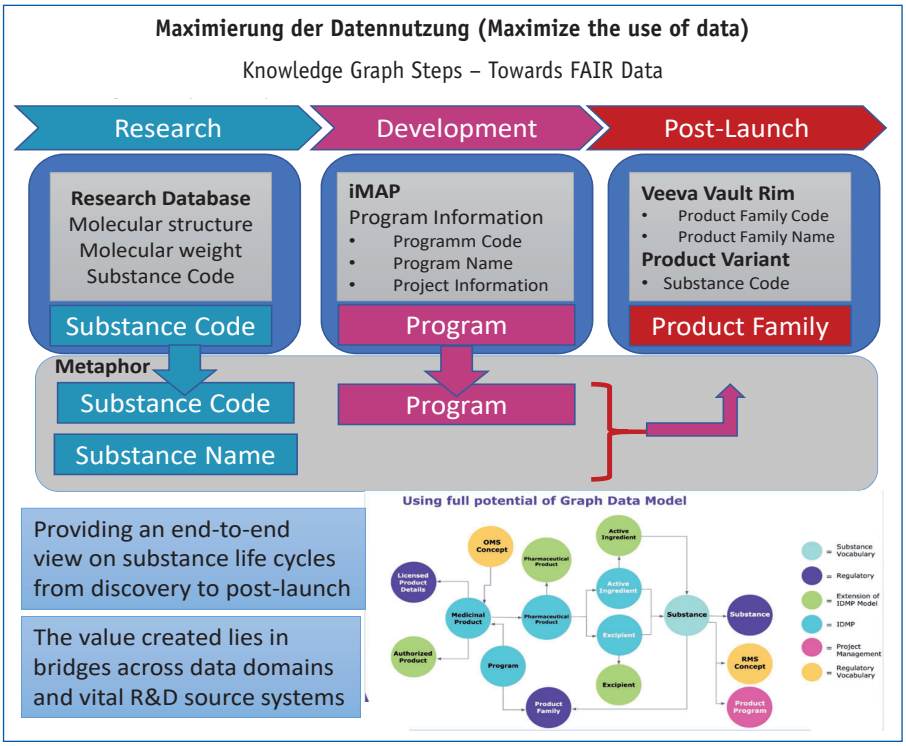


Abb. 3: Maximierung der Datennutzung (Maximize the use of data); Vortrag Merck, BioData World Congress 2021.

sich nicht – falls das nicht bereits zu spät ist – Vertreter der Patienteninitiativen sowie der institutionellen und regulatorischen Institutionen aktiv involvieren, um die politische und ethische Folgenabschätzung, den Diskurs über das Menschen- und Systembild in diesem „neuen Gesundheitswesen“ sowie die Entwicklung von kritischen Implementierungsszenarien einzubringen.

Aber auch die Versorgungsforschung ist gefragt, die weder international noch national auf diesem Weltkongress vertreten war. Dabei wäre die Diskussion um die Kongruenz mit gesellschaftlichen Gesundheitszielen, die Fragen zur Finanzierbarkeit und vor allem der zu erwartende Nutzen eigentlich ein „Must be“ für alle, die mit integrierter agiler Daten- und Wissensbewirtschaftung im Gesundheitswesen befasst sind. Auch wäre eine konzertierte Initiative zur Begleitung dieser Prozesse durch einen Diskurs in der Versorgungsforschung die Kür, wenn nicht gar Pflicht. <<

von: Martin Klein, freier Journalist

* Abkürzungen

SPOR: Substance, Product, Organisation and Referential master data
IDMP: IDentification of Medicinal Products
CTI: Clinical Trials Information System
Elixier Health Data Focus Group
IMI: Innovative Medicines Initiative
Pistoia (nach ital. Stadt): AstraZeneca, GSK, Novartis und Pfizer (https://www.pistoiaalliance.org)
GA4GH: Global Alliance for Genomics and Health (https://www.ga4gh.org)
AAIH: Alliance for Artificial Intelligence in Healthcare (https://www.theaaih.org)
F:A:I:R: Findable, Accessible, Interoperable, and Re-usable (https://www.go-fair.org)

FAIR-Bridging der F&E-Datendomänen von Merck Healthcare

FAIR supporting activities	Steps to FAIR
Connecting R & D Data Domains Interoperability framework defining community practices for data sharing, data formats, metadata standards Adressing data governance ensuring data quality throughout the lifecycle	**Unique Ids** for our vocabularies, representing the actual „thing“. Ensure that all „things“ are always referenced correctly globally unique and never changing
Knowledge representation Sufficient accuracy to utilize information for digital communication being enforced in the next years	**Describing our datasets** being generous, extensive, including descriptive information about the context, quality & condition, or characteristics of the data, including provenance & lineage information
Data Stewardship Supportiing interconnection with data stakeholders primary and secondary use of data	**Semantic modeling** allows us to focus on meaning and knowledge, compared to just describing data structures
Supporting upcoming legislations (IDMP, SPOR, CTIS, EMA-Portals) Enabling data sharing and collaborations across internal data domains, HA`s exchange and companies	**Integrated controlled vocabularies** tie-in to standard vocabularies. These vocabularies exist within our teams, groups, company or external
	Simple Knowledge Organization System (SKOS) captures similarities, makes it explicit and enable data and technology sharing across diverse applications

Tab. 1: FAIR-Bridging der F&E-Datendomänen von Merck Healthcare (FAIR Bridging of Merck Healthcare R&D Data Domains): „Um FAIR zu sein, müssen fünf verschiedene Schritte befolgt werden“. Aus: Vortrag Merck: BioData World Congress 2021.

Wissen ist Macht



Was die internationale Pharma-Großindustrie und ihre vielseitigen IT-Dienstleister in den letzten Jahren auf die Beine gestellt haben, um den Schatz der „Big Data“ heben zu können, ist schon beachtlich. Chapeau! Doch bei aller Hochachtung vor dieser mit enormen Budgets geleisteten Arbeit vieler Know-how-Träger bleibt doch ein fader Nachgeschmack ob des Lobs, der in eine Frage mündet: Warum machen so etwas nicht öffentliche Institutionen, Forschungseinrichtungen und Universitäten? Hier geht es eben nicht um „irgendetwas“, sondern einerseits um wichtige Grundlagenforschung und damit erzielbare Markt- und Wissensvorsprünge – und andererseits um die Gestaltungs- und Definitionshoheit dessen, was schon heute und vor allem künftig mit dem Datenschatz „Big Data“ machbar ist und sein wird. Oder ganz schlicht um die Frage, wer vom „data value continuum“ am meisten profitieren wird: nur die Industrie? Oder doch, so die leise Hoffnung, Patienten und Gesellschaften?