





Gegen alle Widerstände – Daten für die Forschung

Frank Ückert



dkfz.

GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

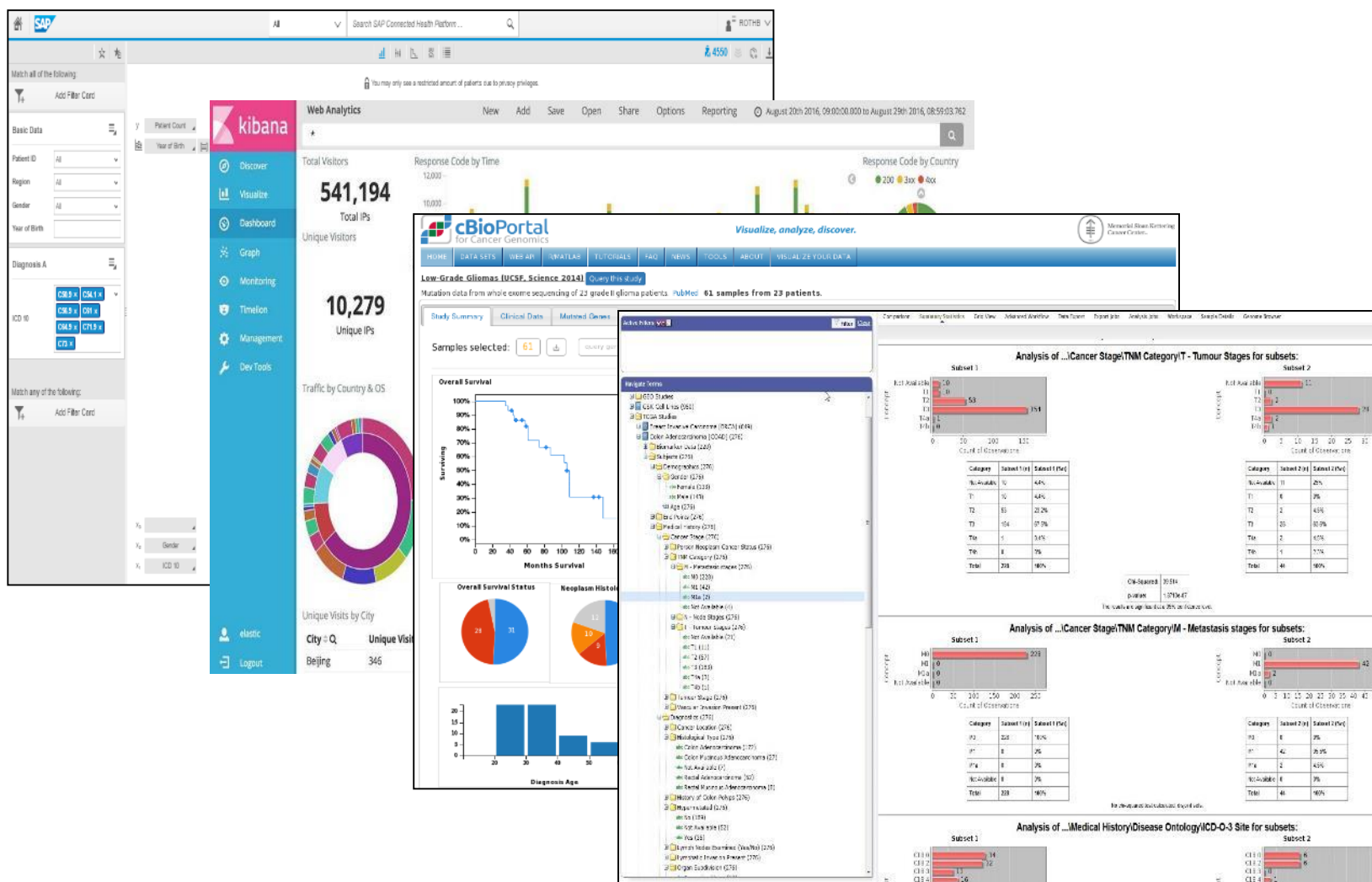


Research for a Life without Cancer

Why hunting for data in the first place?



Immediate benefit: analytic warehousing



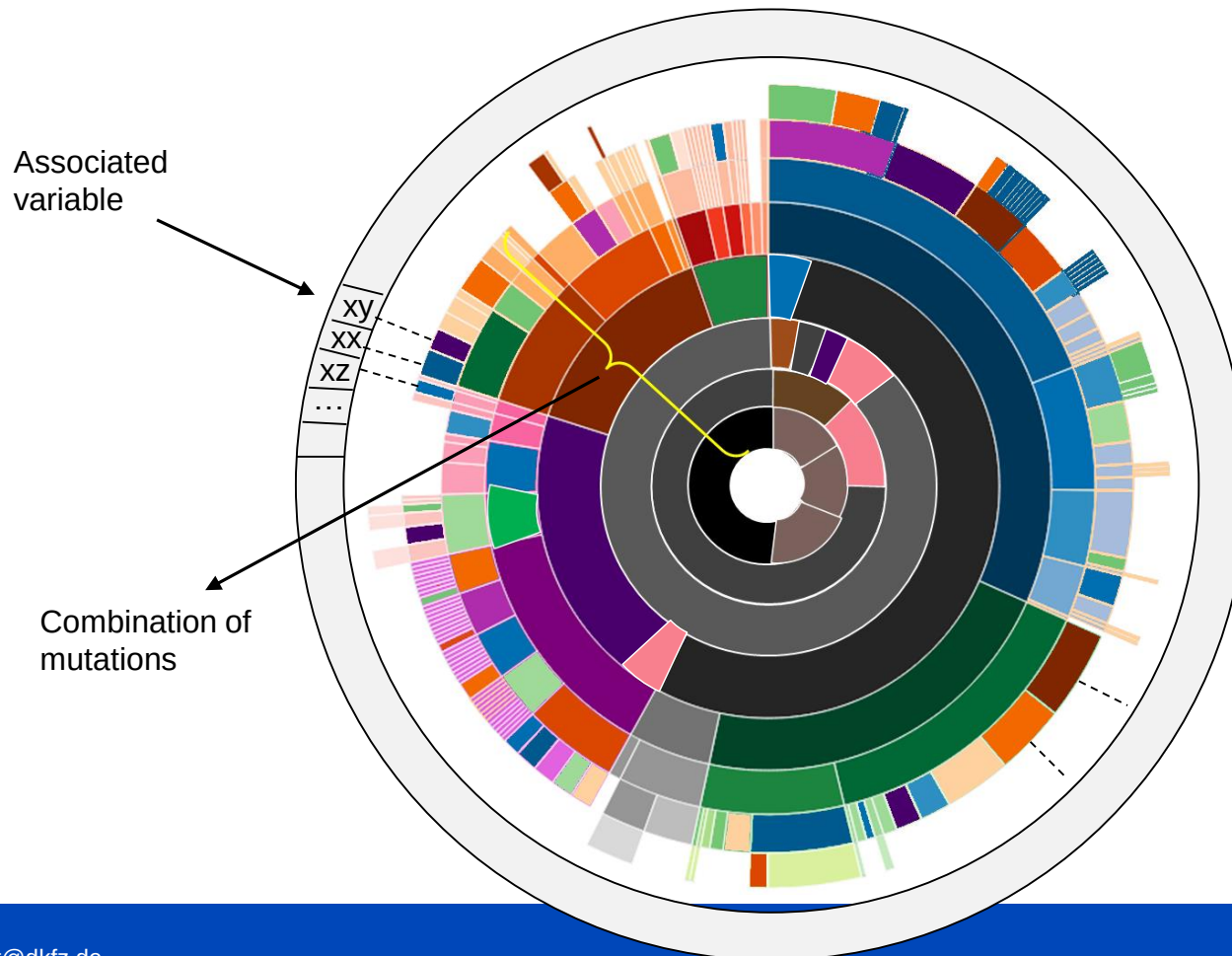
Ganz einfache Fragen...

- Wie viele freigegebene Bioproben aus einem Primärtumor enthält die Biobank von Patienten mit einem Prostatakarzinom und einem Grading der Kategorien mäßig, mäßig gut und mittelgradig?
- Automatisiert ermittelte Knochendichte bei Patientenkohorten mit der Diagnose MGUS (Monoklonale Gammopathie unbestimmter Signifikanz) und mit/ohne anschließende Diagnose Multiples Myelom?
- Wie hoch war die 5-Jahres-Überlebensrate von Patienten mit Diffus-Großzelligem Lymphom, die 2010 im UKHD behandelt wurden?
- Wie lange war die durchschnittliche Zeit zwischen letzter Chemotherapie und Todesdatum bei Tumorpatienten am Beispiel Lymphom?
- Bei wie vielen wurde ein Thorax-CT oder eine Broncho-alveoläre Lavage durchgeführt?
- Wie viele Tumorpatienten hatten eine Pneumonie?
- Wie viele stationäre Tumorpatienten bekamen eine Sepsis?

Aus unserer Vorführung des neuen Datawarehouses im Juli 2017

Immediate benefit: more time of PHD candidates

How does a certain combination of mutations relate to a certain variable of interest (e.g. tumor burden)?



Information visualization of multidimensional data

Using Dashboard Networks to Visualize Multiple Patient Histories: A Design Study on Post-Operative Prostate Cancer

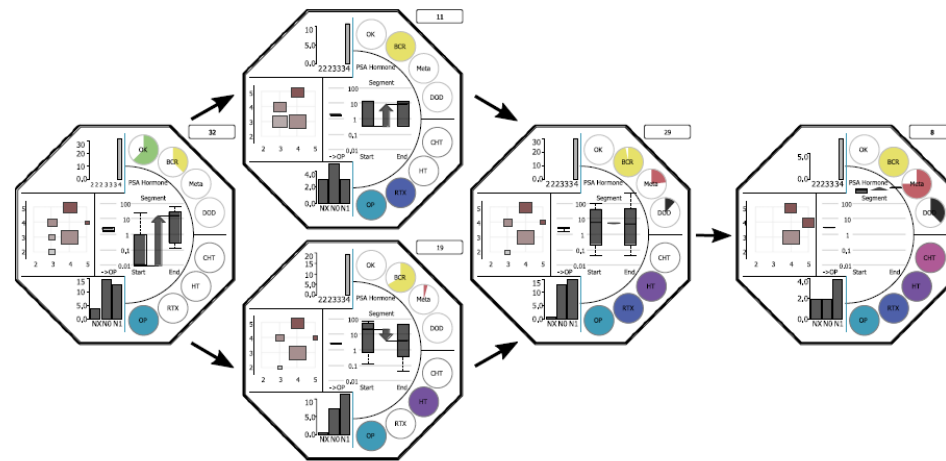


Fig. 3. The history of a cohort of patients who all had a pT-status of pt4. From left to right the static dashboards are ordered by the number of treatments that the patients received (1-4). Two different signatures of treatments exist in the second segment (OP-RTX and OP-HT). Twenty-nine patients have a third segment and eight have a fourth, showing that this cohort contains patients who have the most severe treatment (CHT) and get metastases (Meta) and die (DOD). The colors are the same as shown in Fig. 1.

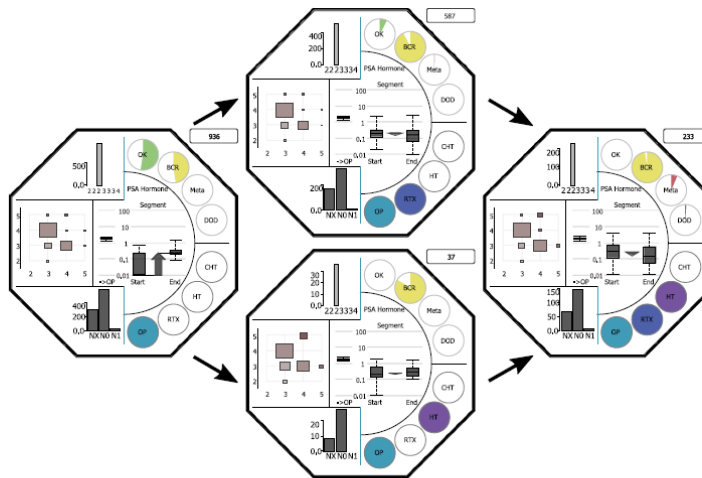


Fig. 4. The history of a cohort of patients who all had a pT-status of pt2c. From left to right the static dashboards are ordered by the number of treatments that the patients received (1-3). Two different signatures of treatments exist in the second segment (OP-RTX and OP-HT). Over nine hundred patients started with segment one and over two hundred have a third segment. However, this cohort contains only few cases with metastases (Meta), virtually no deaths (DOD) and none of the patients received chemotherapy (CHT). The colors are the same as in Fig. 1.

Comparison between cohorts displayed in separate sets of visualizations by using dashboard networks

Figs. 3 and 4 depict two patient cohorts with a different pTstatus (pt4 and pt2c respectively).

One of the most obvious differences between cohorts is the **missing fourth stage in Fig. 4**, indicating that pT-status pt2c does not lead to chemotherapy (CHT).

In stage one both treatment signatures share OP as the first form of treatment, which is predefined in the dataset.

A closer look at the PSA trend reveals that the PSA value of the pt4 cohort increases considerably stronger in the first stage.

In stage two, the treatments received by the pt2c cohort have less implications on the PSA value progression.

One hypothesis to explain this may be that low PSA values are less responsive to treatments.

These observations indicate that pT-status pt2c has a better prognosis than pt4

Visual analytics in science and technology

Clustrophile2: Guided Visual clustering Analysis

Clustering View

An **atomic** representation of clustering instances

- Show variation within clusters
- Allow quick iteration over parameters
- Represent clustering instances compactly
- Support reasoning about clustering instances
- Facilitate interpretable naming



International Investments in AI

- privat and public investments in 2016
 - Europe: €3 billion
- federal budget:
 - European Commission: €1.5 billion by 2020
 - France: €1.5 billion by 2022
 - UK will follow soon



- privat and public investments in 2016
 - Asia: €10 billion
 - US: €18 billion
- federal budget:
 - India: €408 million in 2018-19
 - South Korea: €760 million in 2016-21
 - Canada: €106 million in 2017

German Initiatives in AI

- coalition agreement:
push developments in AI
- €30 million for competence centers for ML
 - Berlin
 - Dortmund/St. Augustin
 - München
 - Tübingen



- Ongoing excellence initiatives:
 - IMAGINE: machine learning in bio-medical imaging and medicine (TUM, HGF)
 - CASCADE : Computational Science for Complex Systems (UBN, HGF)
 - Science of Intelligence (HUB, TUB, MPI)

Oder gerade von gestern:

THEMEN

FORSCHERPORTRAITS

ENERGIEWENDE

KLIMAWANDEL

KREBS

PRESSE & MEDIEN

PRESSEINFORMATIONEN

AKTUELLE

AUSSCHREIBUNGEN

[Home](#) ▶ [Aktuell](#) ▶ [Presseinformationen](#) ▶ [Helmholtz investiert künftig...](#)

Helmholtz investiert künftig 11,4 Millionen Euro pro Jahr in Künstliche Intelligenz und Maschinelles Lernen

Für die Analyse komplexer Daten von der Klima- bis zur Gesundheitsforschung baut die Helmholtz-Gemeinschaft ab sofort sechs Einheiten für Künstliche Intelligenz und Maschinelles Lernen auf. Dazu investiert die größte deutsche Forschungsorganisation dauerhaft 11,4 Millionen Euro jährlich.

„Unsere Gesellschaft befindet sich inmitten eines tiefgreifenden Wandels“, sagt Otmar D. Wiestler, der Präsident der Helmholtz-Gemeinschaft. „Der Klimawandel, die Mobilität der Zukunft oder die Möglichkeiten der Robotik stellen uns alle vor enorme Herausforderungen.“ Künstliche Intelligenz (KI) und Maschinelles Lernen sind mächtige Werkzeuge, um Lösungen für diese großen und komplexen Aufgaben zu entwickeln. Helmholtz wird daher künftig seine breit aufgestellten Kompetenzen im Feld KI noch weiter ausbauen, um gesellschaftlich relevante Fragen zu beantworten. Darüber hinaus treibt Helmholtz die Erforschung und das Verständnis komplexer Systeme voran.

„Dazu hat die Helmholtz-Gemeinschaft eine Forschungsbereichs-übergreifende Struktur geschaffen, in der Kompetenzen in KI und Maschinellem Lernen durch Zusammenarbeit geteilt, erweitert und auf eine neue Ebene gehoben werden können“, so Wiestler weiter. Dieses Vorhaben wird bundesweit mit einer Helmholtz-

Kontakt

Kommunikation und Außenbeziehungen

Helmholtz-Geschäftsstelle

 Geschäftsstelle Berlin
Anna-Louisa-Karsch-Straße 2
10178 Berlin

 +49 30 206329-57

 [presse\(at\)helmholtz.de](mailto:presse(at)helmholtz.de)

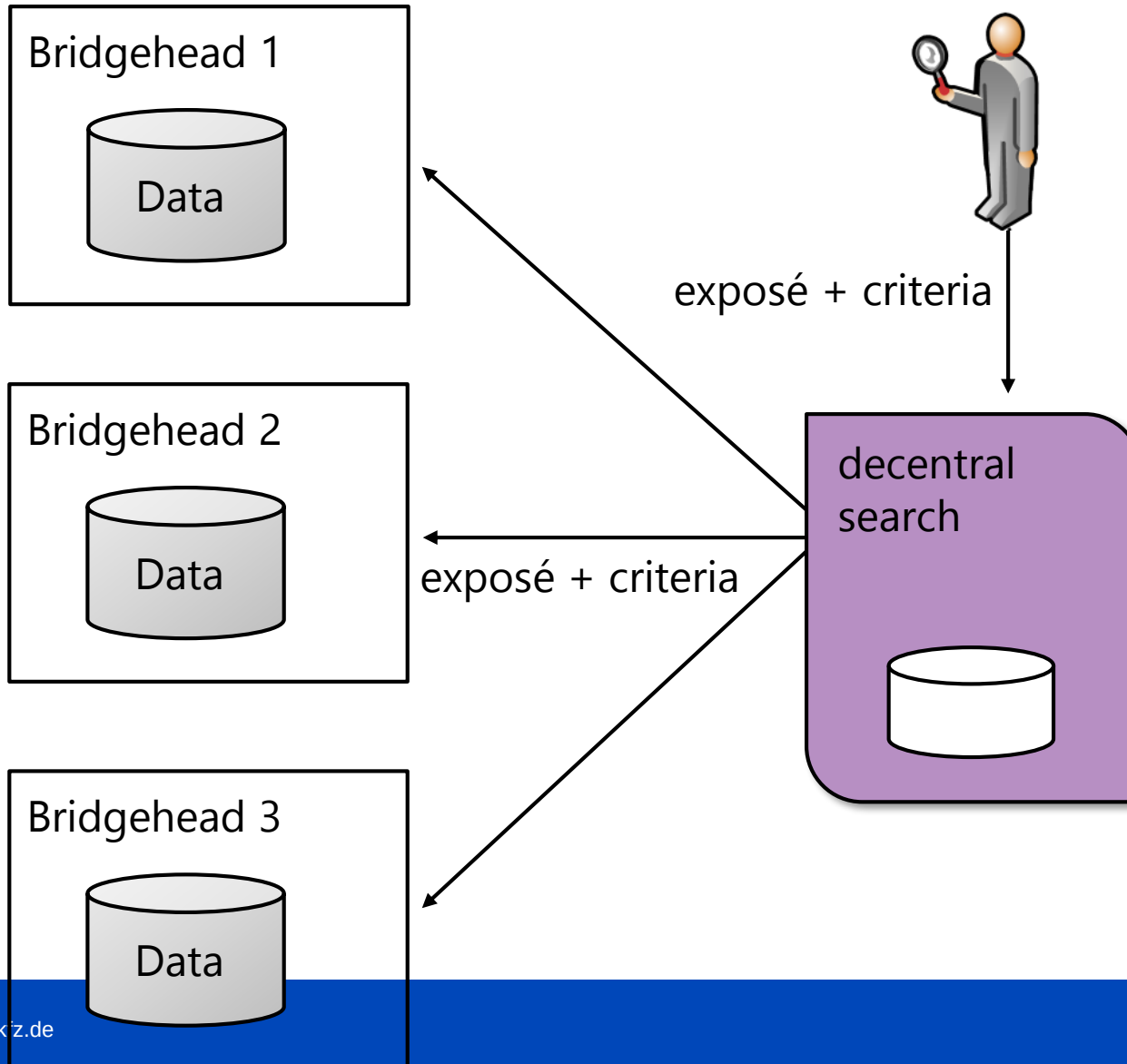
Presseinformationen regelmäßig erhalten

Sie möchten unsere
Presseinformationen per E-Mail
erhalten? Schreiben Sie eine kurze E-
Mail an [presse\(at\)helmholtz.de](mailto:presse(at)helmholtz.de) und
wir nehmen Sie in unseren Verteiler.
Sie können auch unsere ▶
[Presseinformationen per RSS-Feed](#)

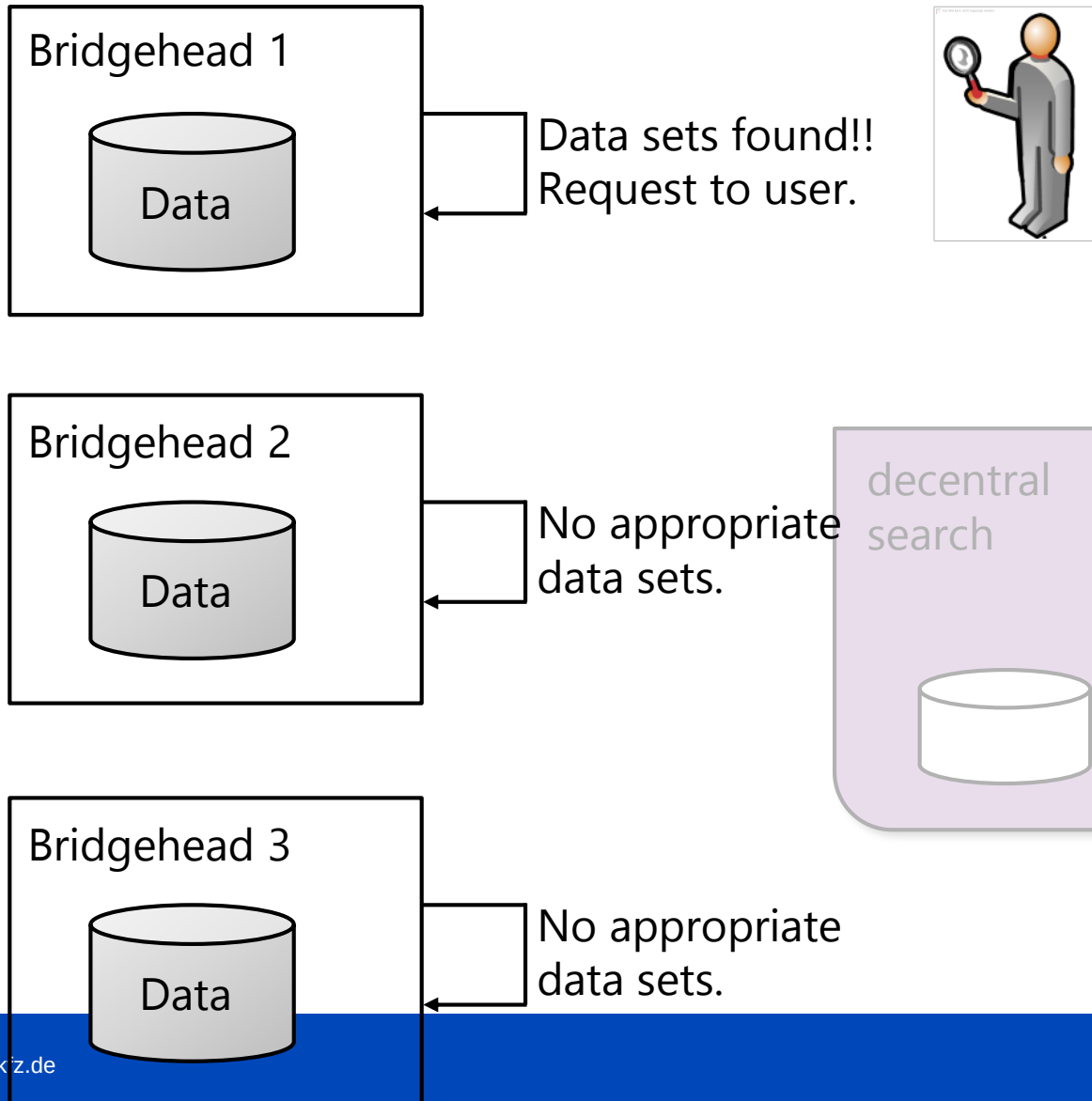
Dezentral ist das Paradies! ... ?



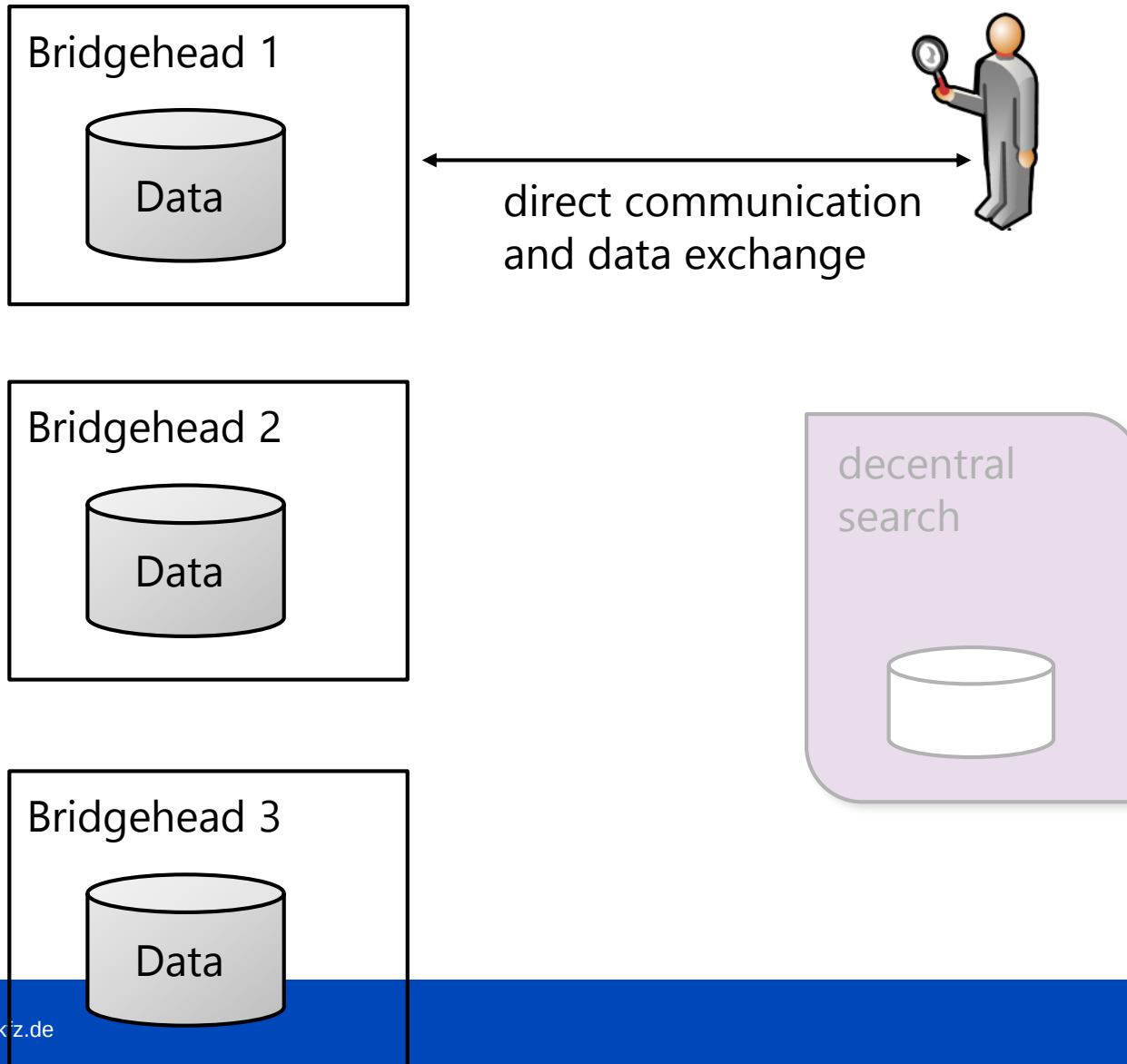
Decentral Search (1/3)

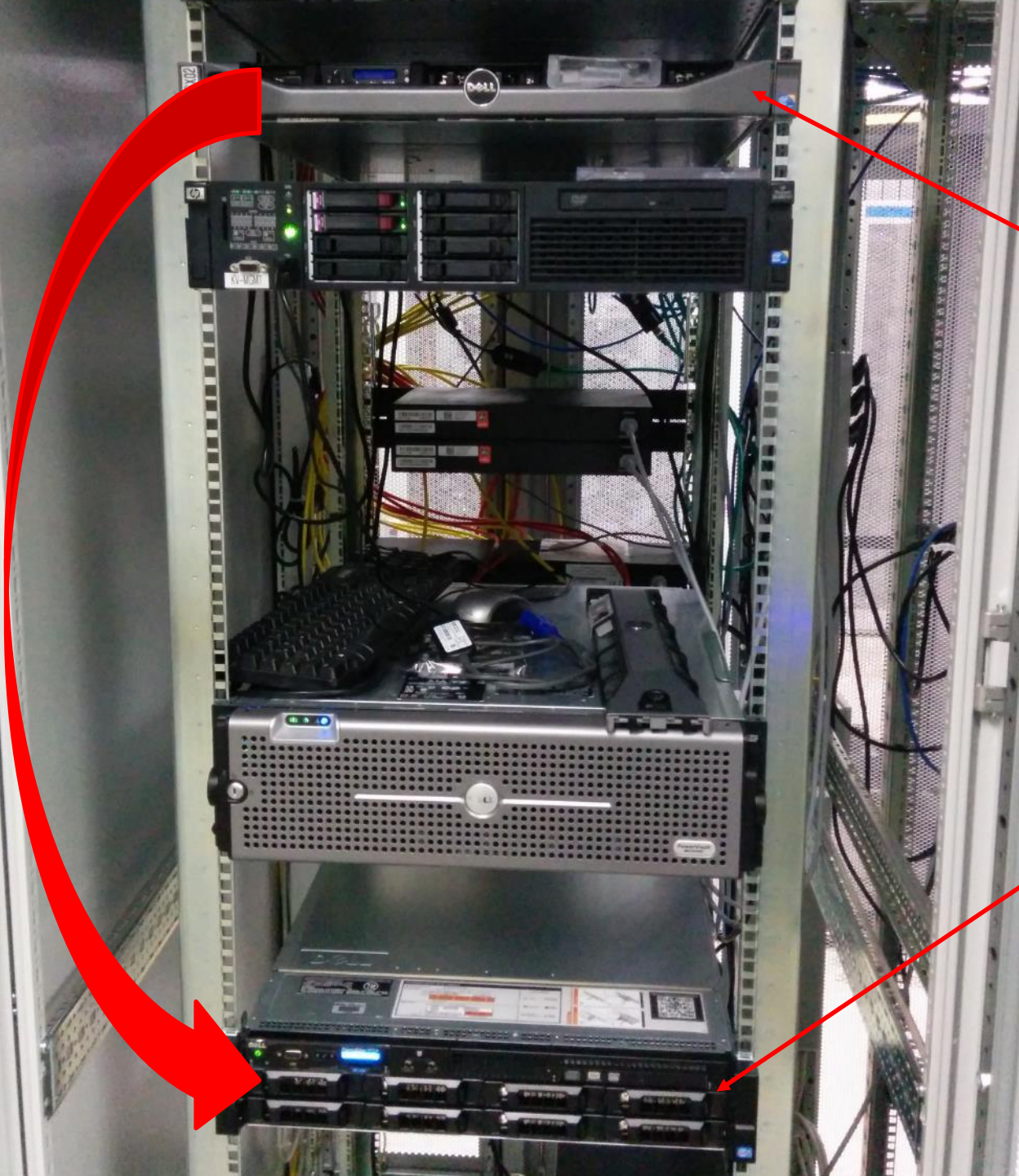


Decentral Search (2/3)



Decentral Search (3/3)

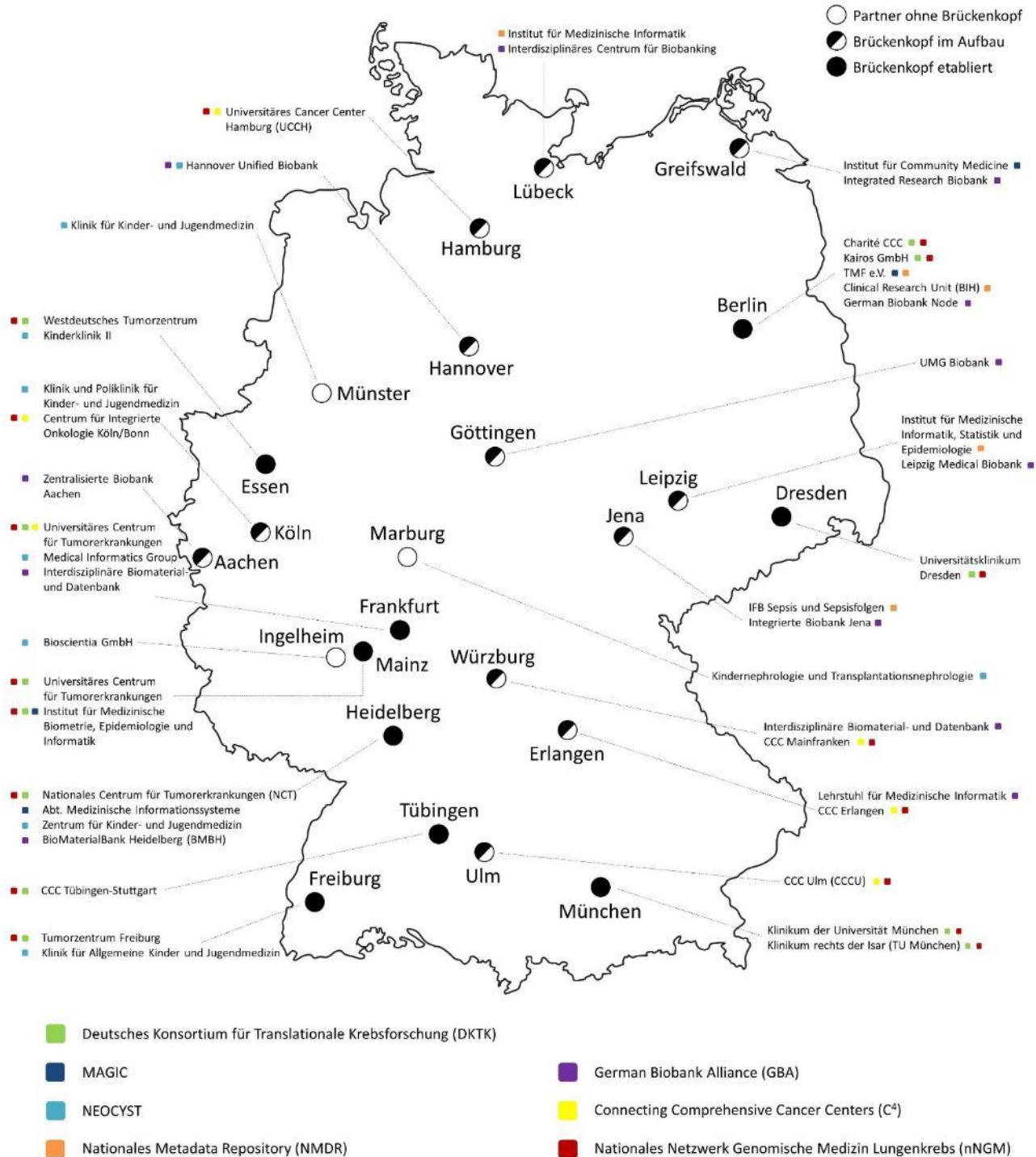




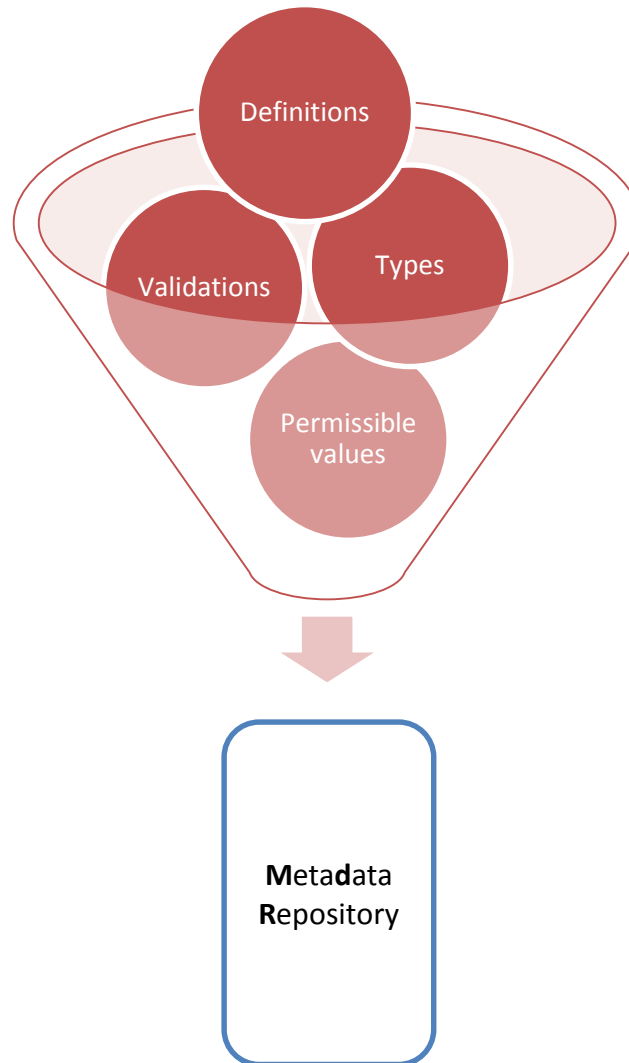
Source System

Bridgehead

Brückenköpfe an Partnerstandorten



One possible show stopper: semantics

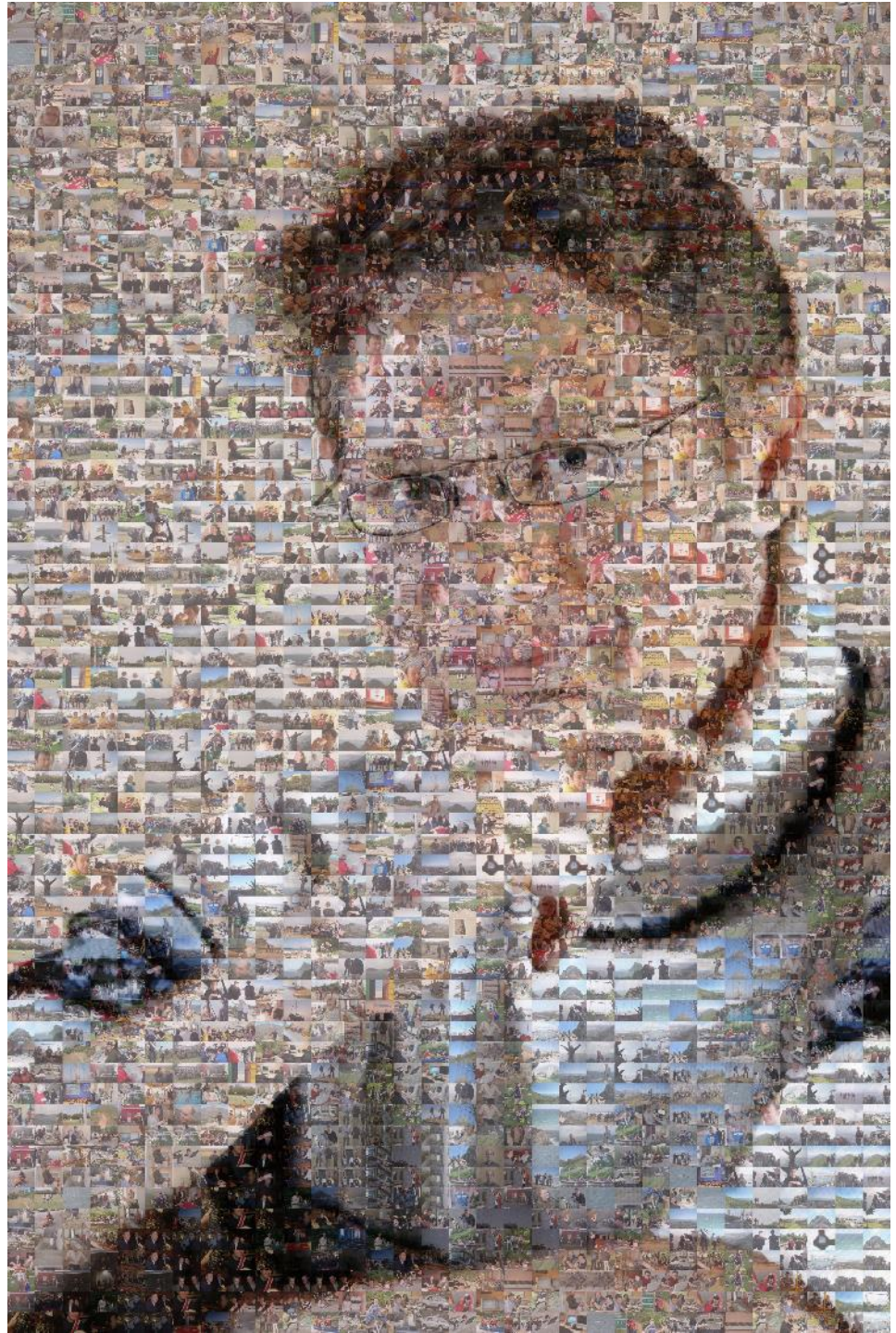


Warum dann überhaupt noch zentrale Datensammlung?

Surplus value comes from integration of heterogeneous data types of the same patient.

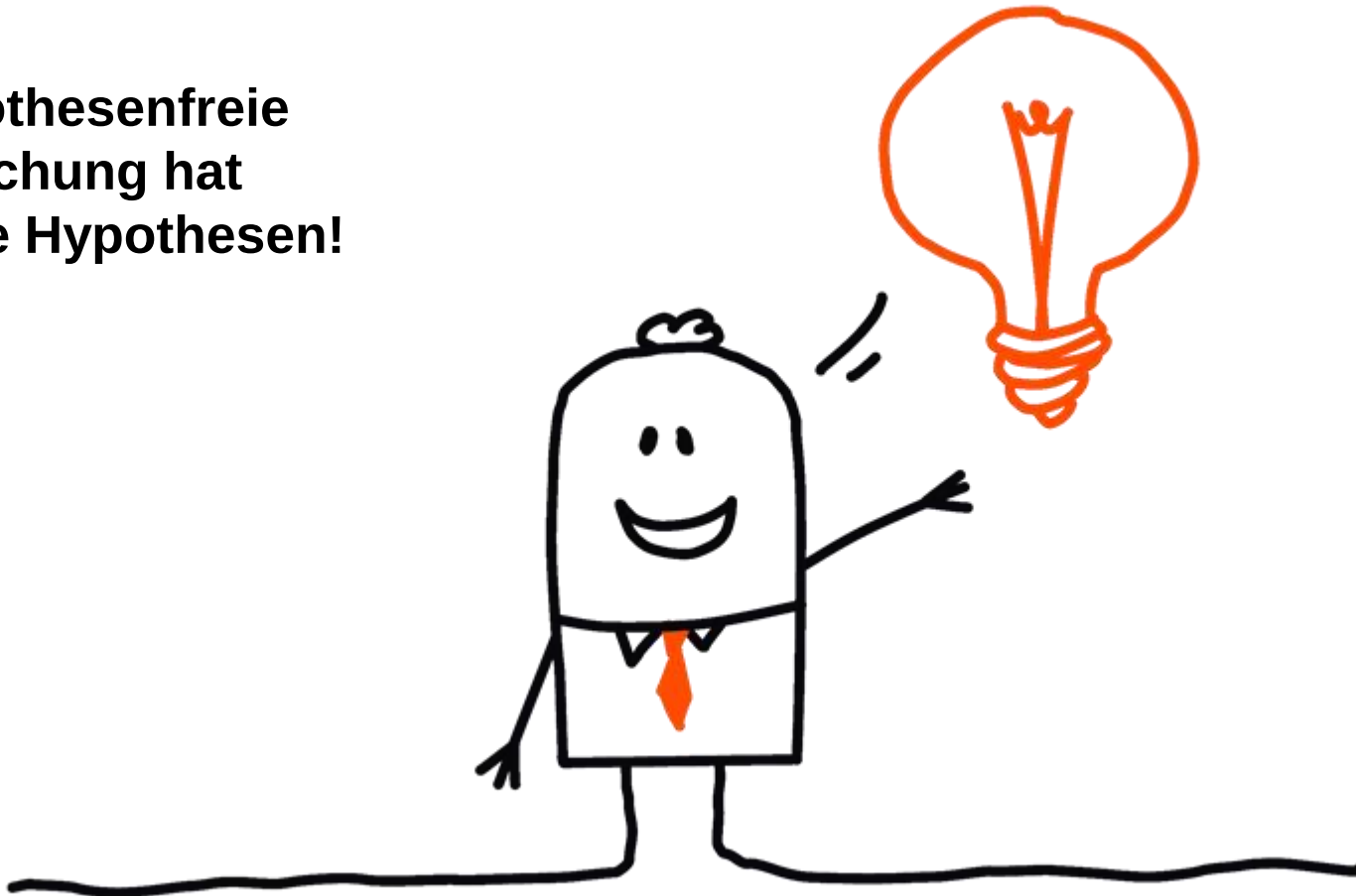


To see the whole picture, the data has to be just there and not be distributed.



Völlig neue Erkenntnis zum Mitnehmen

**Hypothesenfreie
Forschung hat
keine Hypothesen!**



The Holy Grail ...

... is science
and not the scientist!



Often heard

„Only I can make sense out of my data“

Please be part of the group.

„You need additional knowledge/data to interpret the data“

Please invite other partners.

„I keep it exclusively on my external HDD to be safe“

Please consider logic.

„Do not ask me - I am only the guardian“

And that is exactly the person to ask.

„I do not know whom to ask for permission“

Please describe the process and we will find someone..

„It is forbidden“

Let us look into the contracts and perhaps work on a better contract for the next time.

„It is not consented for“

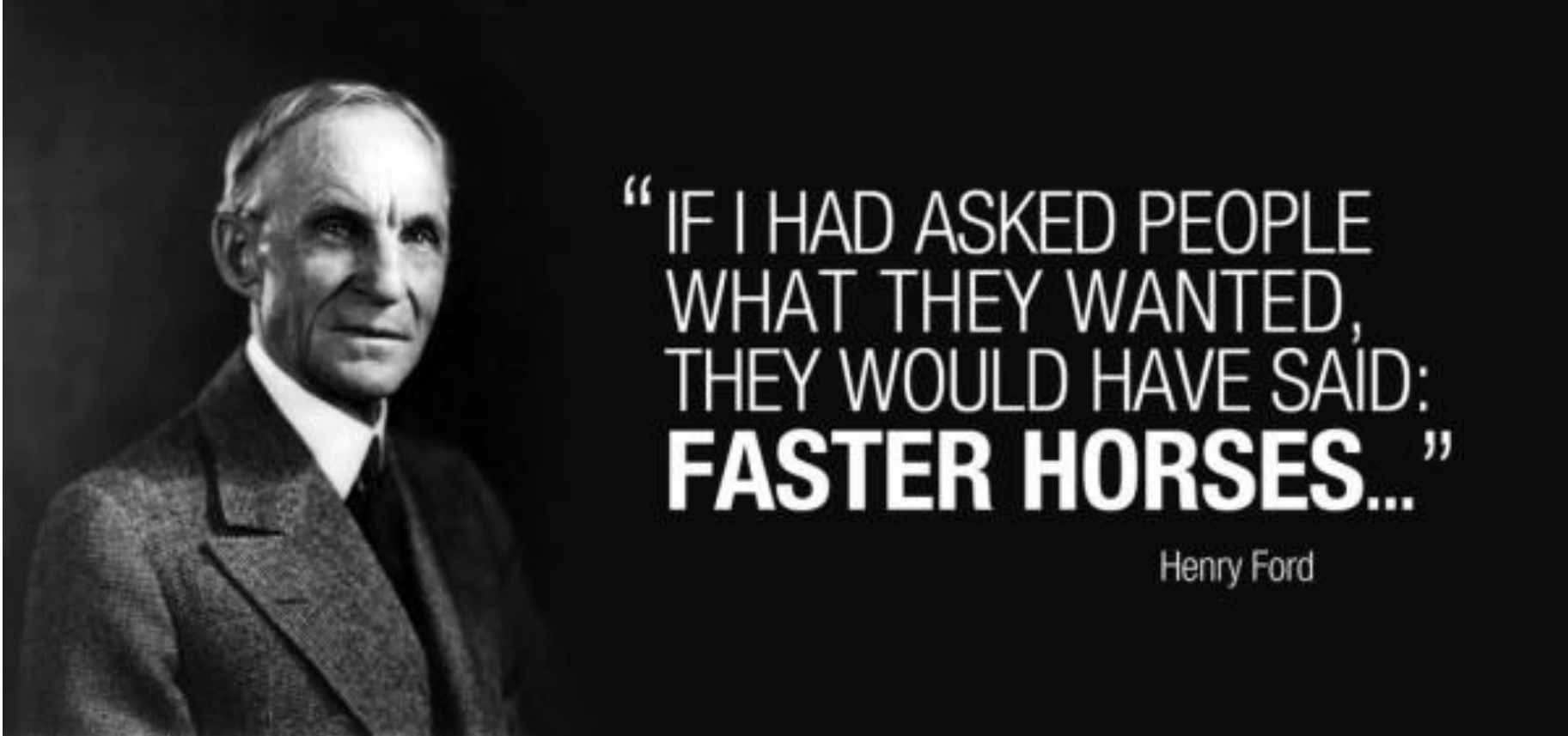
Let us ask the ethical review board and please use the worked out consent form next time.

„Data Protection“

Yawn... So easy, wait a minute.

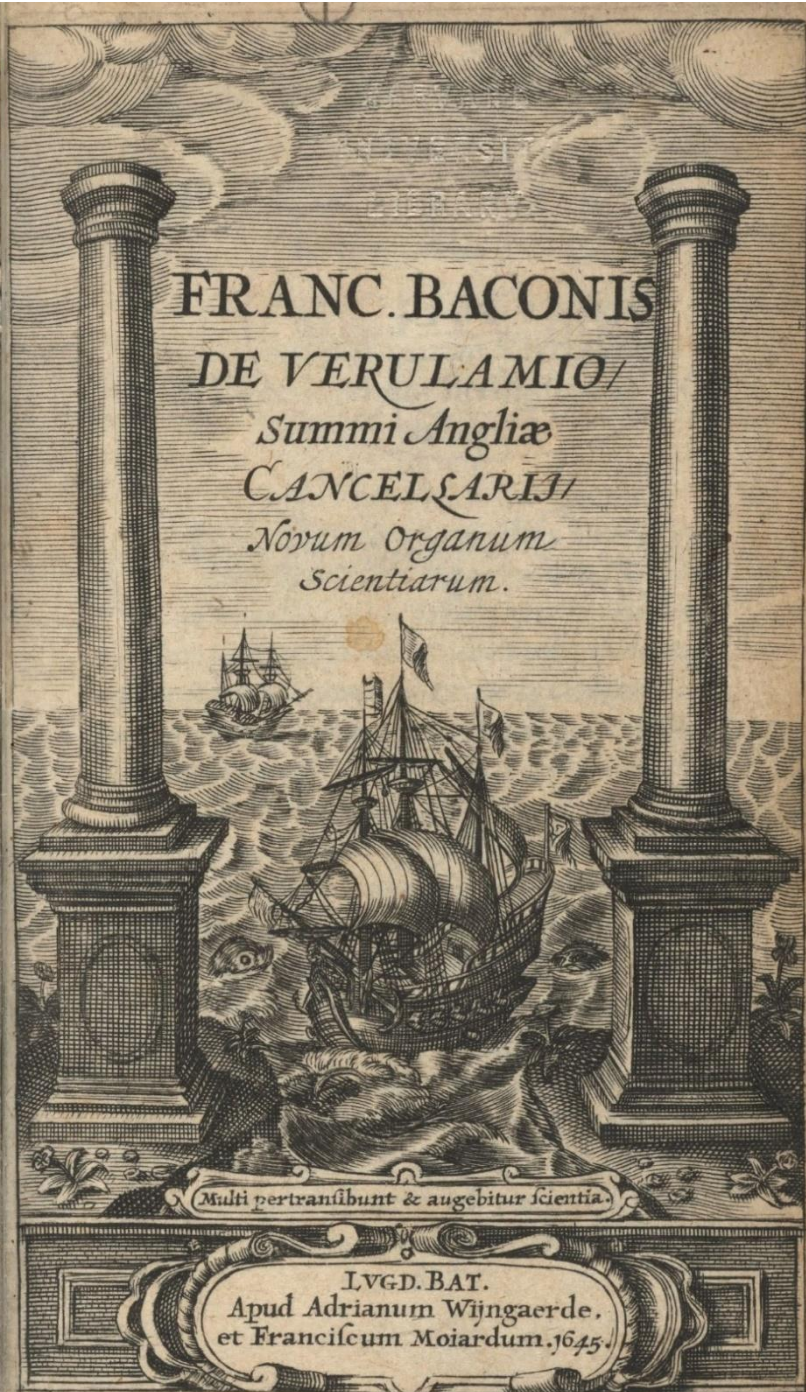
Rarely heard

„I am not interested“





Generating hypotheses
automatically
(with Artificial Intelligence)



Francis Bacon, Novum
Organum, 1645

Scientist working with data (only)

- Experts of different domains
- retreat together
- to generate hypotheses
- in limited, but reserved time
- in stimulating environments
- on present data.





Nationale Dekade gegen Krebs



Mission on Cancer
1,2 bn€

Inevitable alternative





„Die kaufe ich einfach bei der UK-Biobank!“

- Aged 40-69 at baseline assessment during 2007-2010
- Collected at Assessment Centre: 500,000 participants
- Collected at Repeat Assessment Center: 20,000 participants
- Imaging study Data for 100,000 participants
- Genetics Data, Genome-wide 488,000 participants
- Linked health-related Data

„Die kaufe ich einfach bei der UK-Biobank!“



Summary of Basket 2002271

100315 Genomics - Genotypes **Show +26**

100092 Health-related outcomes - Cancer register **Hide -8**

40009 Reported occurrences of cancer

40005 Date of cancer diagnosis

40008 Age at cancer diagnosis (filter)

40006 Type of cancer: ICD10

40013 Type of cancer: ICD9

40011 Histology of cancer tumour

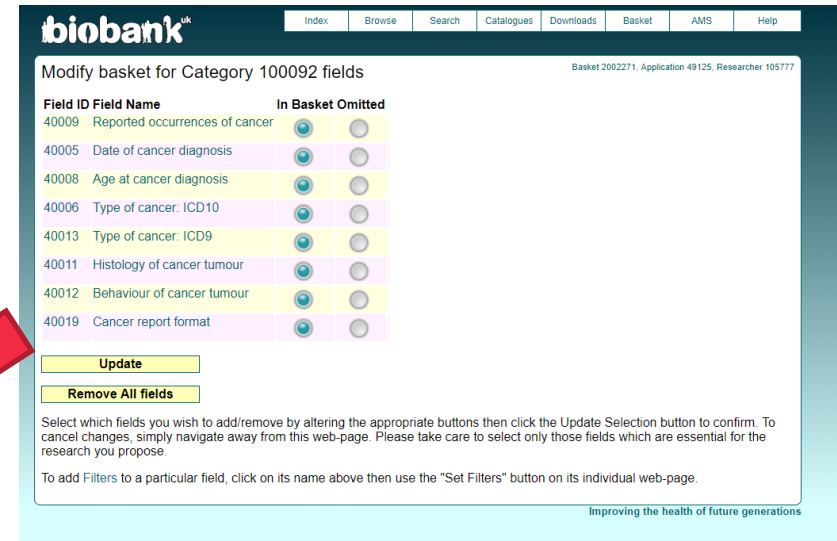
40012 Behaviour of cancer tumour

40019 Cancer report format

Filter Conditions **Previous Baskets** **List Actions**

There is 1 field with filters defined (click name to edit)

40008 Age at cancer diagnosis min=30



biobank^{uk}

Index Browse Search Catalogues Downloads Basket AMS Help

Modify basket for Category 100092 fields

Basket 2002271, Application 49125, Researcher 105777

Field ID	Field Name	In Basket	Omitted
40009	Reported occurrences of cancer	☒	☐
40005	Date of cancer diagnosis	☒	☐
40008	Age at cancer diagnosis	☒	☐
40006	Type of cancer: ICD10	☒	☐
40013	Type of cancer: ICD9	☒	☐
40011	Histology of cancer tumour	☒	☐
40012	Behaviour of cancer tumour	☒	☐
40019	Cancer report format	☒	☐

Update

Remove All fields

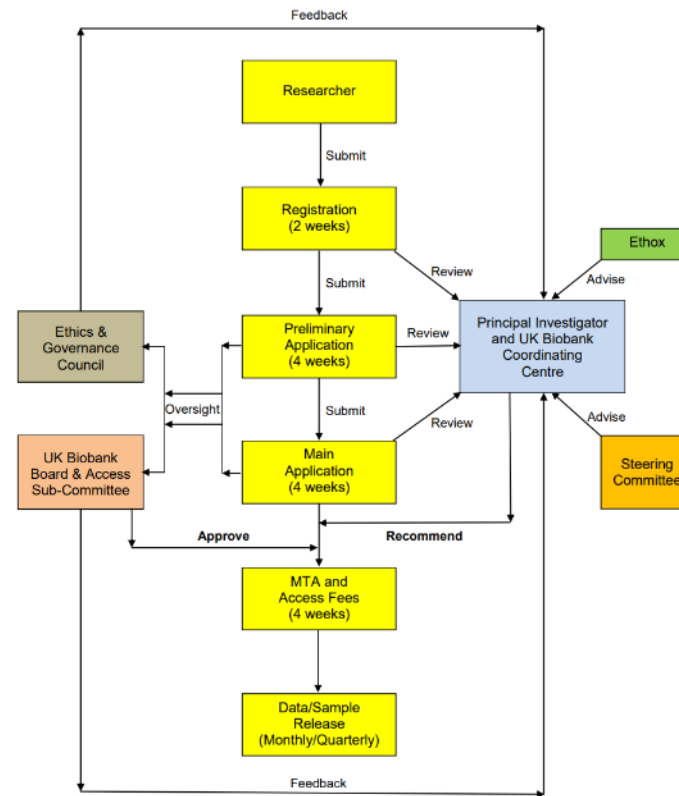
Select which fields you wish to add/remove by altering the appropriate buttons then click the Update Selection button to confirm. To cancel changes, simply navigate away from this web-page. Please take care to select only those fields which are essential for the research you propose.

To add Filters to a particular field, click on its name above then use the "Set Filters" button on its individual web-page.

Improving the health of future generations

“Please give me all data in the cancer register for participants that are at least 30 years old and the corresponding genotypes.”

„Die kaufe ich einfach bei der UK-Biobank?“



Stages in the application and review process (with the indicative timelines in parentheses); the roles of the different parties are described in Section C11

„Die kaufe ich einfach bei der UK-Biobank?“

- A1. Project title (200 characters):
- A2. Research question(s) and aim(s) (up to 5000 characters or 200 words):
- A3. The background and scientific rationale of the proposed research project in general (up to 5000 characters or 300 words):
- A4. A brief description of the method(s) to be used (up to 5000 characters or 300 words):
- A5. The type and size of dataset required (e.g., case-control subset, men only, imaging data only, whole cohort, etc.) (up to 5000 characters or 100 words):
- A6. The expected value of the research (taking into account the public interest requirement) (up to 5000 characters or 100 words):
- A7. Please provide up to 6 keywords which best summarise your proposed research project:
- A8. Please provide a lay summary of your research project in plain English, stating the aims, scientific rationale, project duration and public health impact (up to 5000 characters or 400 words):
- A9. Will the research project result in the generation of any new data fields derived from existing complex datasets, such as imaging, accelerometry, electrocardiographic, linked healthcare data, etc, which might be of significant utility to other researchers:
- A10. What is the estimated duration of your project, in months?

„Die kaufe ich einfach ...“

- C5.1.1 Lay summary (updated, if necessary, from the Preliminary Application);
- C5.1.2 Names of all collaborators (who also need to be approved researchers);
- C5.1.3 Scientific rationale of project (background and any pilot data; experimental details and design; power calculations; expected value of results; relevant references);
- C5.1.4 Required data and/or quantity and type of samples, and any need for re-contact, with a brief justification (updated, if necessary, from the Preliminary Application);
- C5.1.5 Protocols for storage of the data and/or samples (data security; physical security of the samples; ability to handle withdrawals);
- C5.1.6 Feasibility issues (e.g. research capabilities; collaborators);
- C5.1.7 Proposed timetable (start; duration; availability of results; submission of publication);
- C5.1.8 Details of funding (or applications for funding);
- C5.1.9 Details of any peer review (actual or proposed); and
- C5.1.10 Applicants may (but are not required) to submit letters of support from third parties.

6.2 The Applicant shall use its best endeavours to publish the Results within 6 months after the Completion Date for the Research Project:

6.2.1 in an academic journal; or

6.2.2 on an open source publication site.

6.3 By the earlier of 6 months after the publication of the Results, or 12 months after the Completion Date, the Applicant will provide to UK Biobank a copy of:

6.3.1 the Results themselves in such form and format as UK Biobank will reasonably require;

6.3.2 the Results Data in such form and format as UK Biobank will reasonably require.

6.4 UK Biobank will consider reasonably any written requests (containing an appropriate explanation) for an extension of the time limits set out in this clause.



Telematikrahmenarchitektur

Das Gesetz zur Modernisierung der gesetzlichen Krankenversicherung vom 14. November 2003 schrieb die Einführung der eGK zum 1. Januar 2006 in § 291a SGB V gesetzlich fest.

Toll!



Interesting thought: why not fresh from the source?



Gesundheit gemeinsam neu gestalten



Viviv

Elektronische Gesundheitsakte (eGA) und persönliche Gesundheitsassistentin als App:

Unterstützt Sie dabei, medizinische Daten zu bekommen, zu verstehen und zu nutzen.

- › Alle medizinischen Unterlagen gebündelt und jederzeit zur Hand
- › Anfragen von Dokumenten behandelnder Ärzte und Freigabe
- › Gesundheitscheck mit Errechnung des biologischen Alters

viviv.barmenia.de



Viviv



CE

Den Telearzt immer dabei

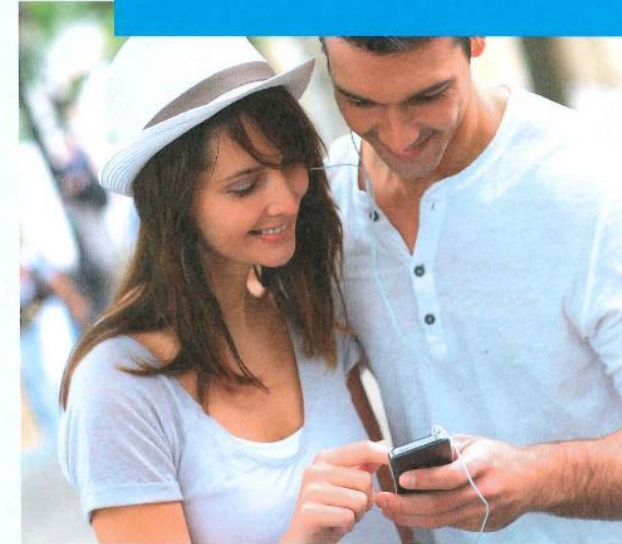
Telemedizinischer Service: Gewährleistet einen schnellen, digitalen Zugang zu qualitativ hochwertiger ärztlicher Beratung.

- › Unabhängig von Öffnungszeiten:
24 Stunden, 7 Tage in der Woche erreichbar
- › Über 200 Ärzte aus über 30 Fachrichtungen
- › Zugang zu erstklassigen Medizinern unabhängig vom Wohnort
- › Erreichbarkeit von deutschsprachigen Ärzten bei Reisen im Ausland
- › Zweitmeinung

mediapp.barmenia.de



**Barmenia
MediApp**



Digitale Gesundheitsservices

mySugr: Diabetes Management einfach und digital

- › Mit der mySugr App und dem Blutzuckermessgerät jederzeit alle Daten zur Hand

Online-Unterstützungsprogramme bei psychischen Erkrankungen

- › Hilfestellung und Beratung bei psychischen Erkrankungen
- › Übungen, Methoden und Techniken zur Krankheitsbewältigung
- › Rückfallprävention u.v.m.

digitale-services.barmenia.de



Gesundheitsakten, die Ersten

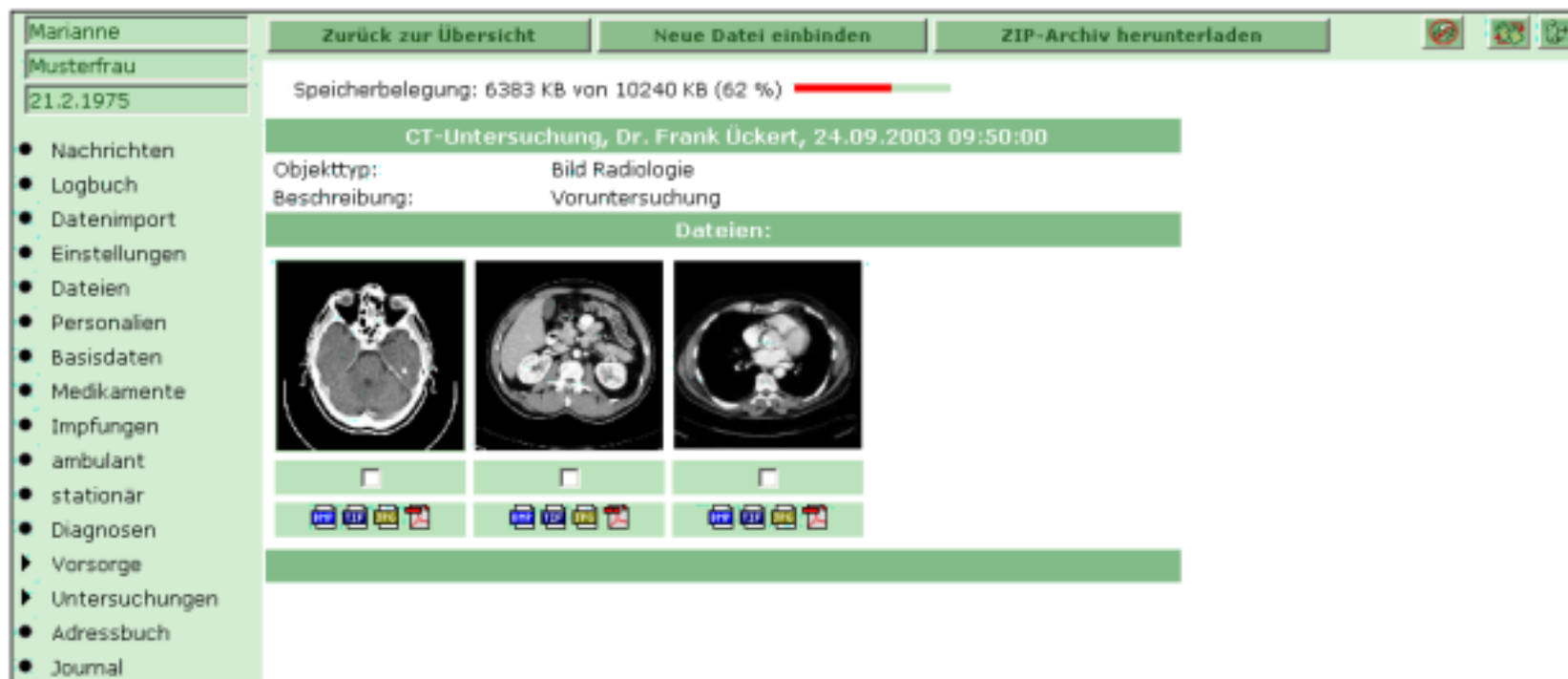


Figure 2: Detailed view of a multimedia data object (in this example DICOM radiology images) with thumbnail preview within akteonline.de.

Gesundheitsakten, die Ersten

akteonline.de

Meine Gesundheitsakte

[Akte wechseln](#)
[abmelden](#)

meine Akte

Mustermann, Karl

Hauptmenü

- meine Daten
- meine Medikamente
- medizinisches Tagebuch
- Diagnosen
- Arztbesuche
- Untersuchungen
- Impfungen
- Vorsorge

Optionen

- Sicherheitseinstellungen
- Logbuch
- Schnittstellen / Dateien

Selbsthilfe

- Forum
- Journal

Neues Medikament
<< 10 zurück 10 weiter >> [Mehr anzeigen](#) ?

Medikament	Wirkstoff	Darreichungsform	Dosierung	
▼ Zyrtec	Cetirizin	Tabletten	0 - 0 - 1	
▼ Belok-Zok mite	Metoprolol	Tabletten	1 - 0 - 0	
▲ Aspirin plus C	Acetylsalizylsäure	Tabletten	bei Bedarf	
Einnahmezeitraum		seit Oktober 2003		
Grund der Einnahme		Kopfschmerzen		
Verschreiber				
Individuelle Wirkung		Nachlassen der Kopfschmerzen		
Patienteninformationen zu 'Aspirin'				
▼ Advantan Creme	Methyprednisolon	Creme	2x tägl.	

Gesundheitsakten, die Ersten

careon Gesundheitsakte - Windows Internet Explorer

careon Gesundheitsakte

»Einstellungen »Alte drucken »Alte schließen

 Gesundheitsakte

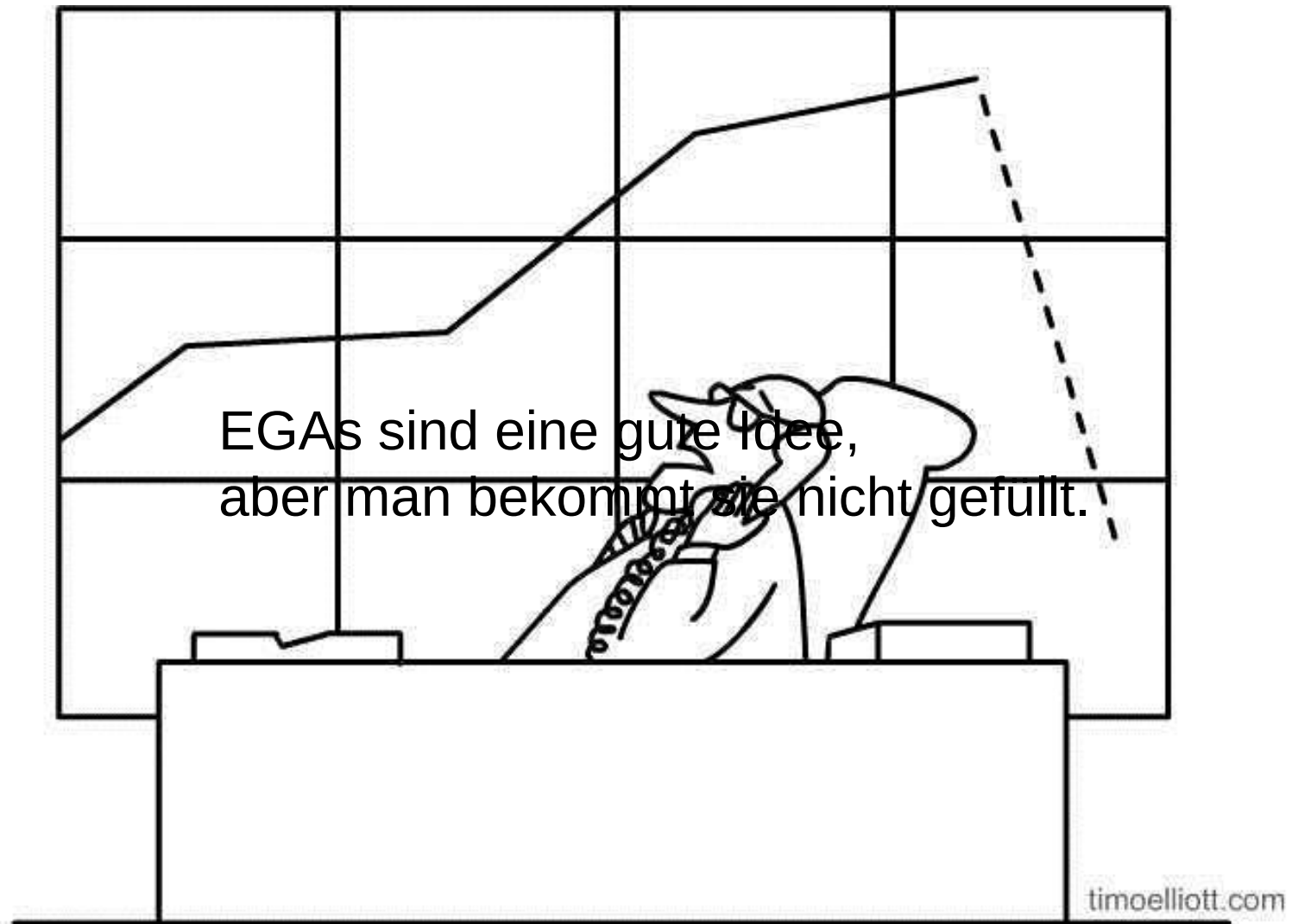
Gesundheitsakte von Max Mustermann

Elektronische Patientenquittung

Alle Leistungsbereiche

Datum von/bis	Kurzbeschreibung der Leistung	Betrag	Zu zahlung	Be zahlung	
09.07.2007	ARZNEI- UND HILFSMITTEL (APOTHEKE) 1 x CLINDAMYCIN RATIO 600MG Apothete: (Buckel) Apotheke Eidelstedt Hoch, Dinstaten Verordnender Arzt: »Münzen, Rita	16,35 €	0,00 €	☒	☐
11.12.2006 - 12.12.2006	AMBULANTE ARZTLICHE LEISTUNGEN »Volbrandt, Ulrich Orthopädie	46,07 €		☒	☐
08.08.2006 - 10.08.2006	KRANKENHAUSLEISTUNGEN Augusta-Hirschen-Anstalt Ev.-Krankenhaus	706,36 €	20,00 €	☒	☐
03.06.2006	ARZNEI- UND HILFSMITTEL (APOTHEKE) 1 x DECODORIN TRI Apothete: Die Maßener-Apothete Dieter Reiss, Dinstaten Verordnender Arzt: »Münzen, Rita	20,38 €	5,00 €	☒	☐
27.04.2006	HILFSMITTEL 1.00 x Weichspoleneinlagen 2.00 x Formabdruck in eigener Werkstatt, incl. Postle 1.00 x Weichspoleneinlagen Bergler Thomas Verordnender Arzt: »Peterski, Peter	119,52 €	11,96 €	☒	☐

Internet 300%



"It turns out that failure WAS an option after all..."

DataBox-Projekt: Digitales Gesundheits-Management in Patientenhand



Forschen für ein Leben ohne Krebs



Netzwerk
Genomische Medizin
Lungenkrebs



nNGM

Nationales Netzwerk
Genomische Medizin
Lungenkrebs



GEFÖRDERT VOM

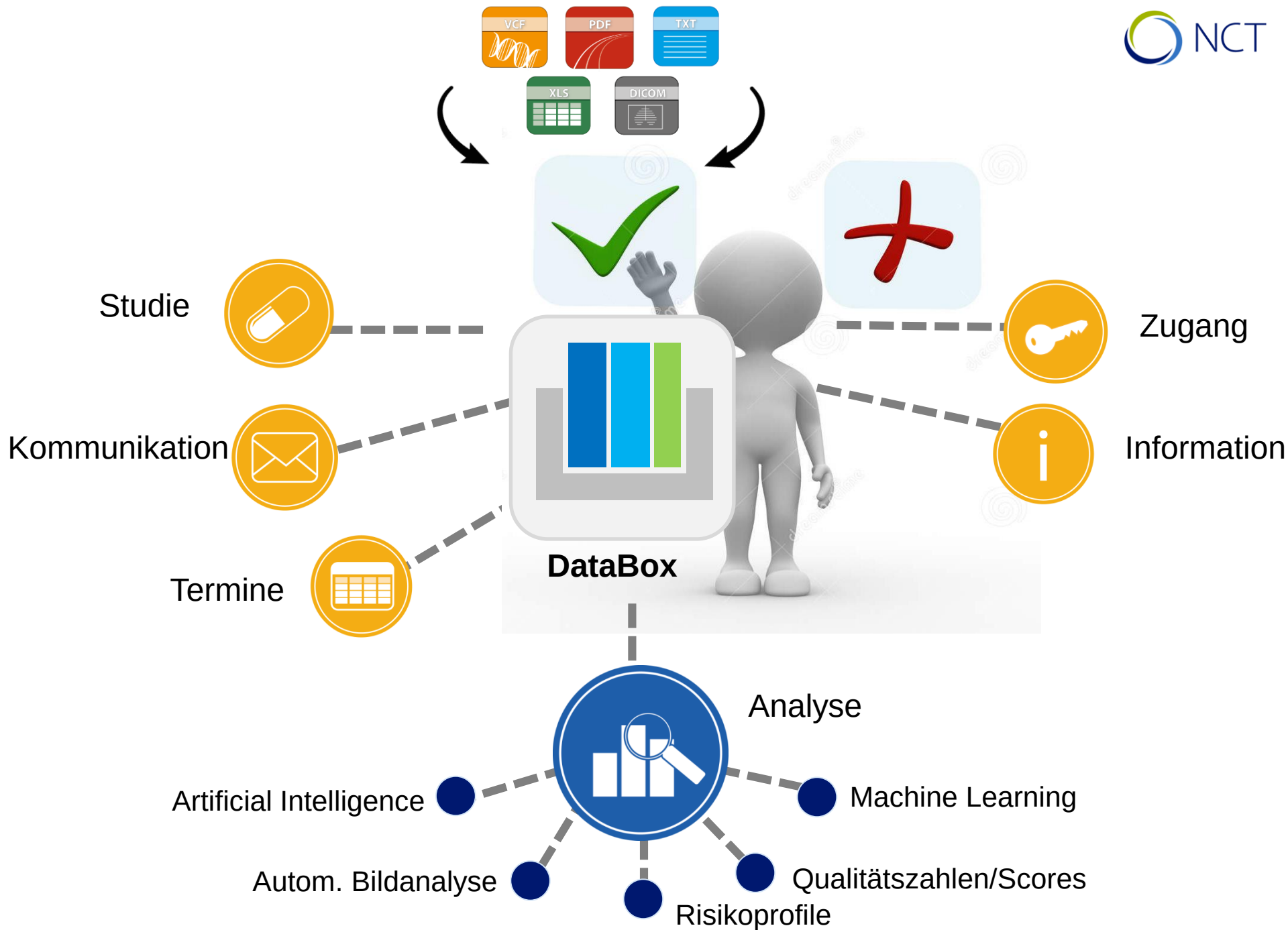
Bundesministerium
für Bildung
und Forschung



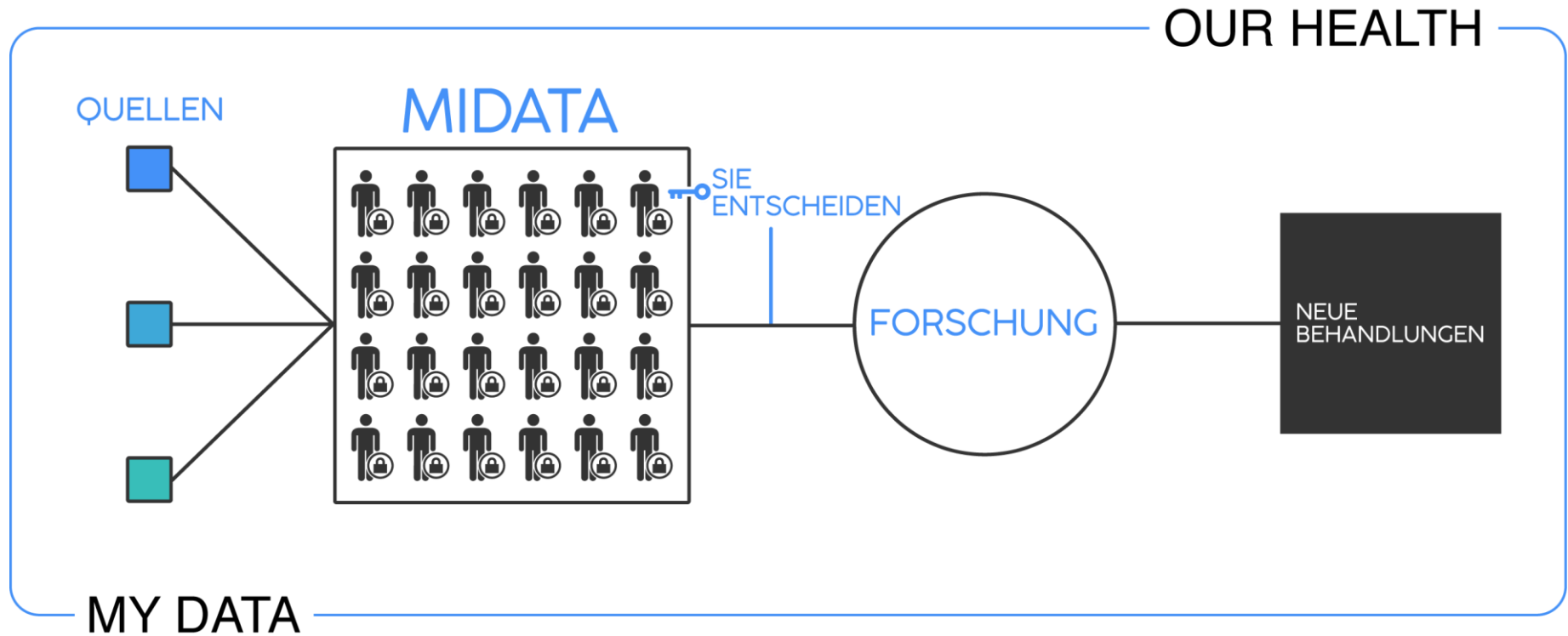
Gefördert durch:

Bundesministerium
für Gesundheit

aufgrund eines Beschlusses
des Deutschen Bundestages



Genossenschaft!



Resümee



A good solution



Lessons Learned

1. Aim for „good“ data only (and be true to yourself)
2. Aim for sharp data (instead of broad data)
3. Go and search for data (e.g. by searching for and following interests of third party stakeholder)
4. Agree on a core data set as „internal open data“ (one per new data collection)
5. Establish a small support unit (mainly for overcoming ELSI problems)
6. Build trust (e.g. via governance methods like veto rights, personal data sheriff)
7. Give immediate benefits (which is easy because of HIS)

What was in it for me personally?





NATIONALES CENTRUM
FÜR TUMORERKRANKUNGEN
HEIDELBERG



UniversitätsKlinikum Heidelberg

dkfz.

DEUTSCHES
KREBSFORSCHUNGSZENTRUM
IN DER HELMHOLTZ-GEMEINSCHAFT



Forschen für ein Leben ohne Krebs