Unit 3 - Basics of Information Management

Definitions and conceptual demarcation

1.1 Information management

Data

Data is originally a plural, the rarely used singular is datum. Data is measured and documented in different structures, formats and units. This depends on the type of data and also very much on its subsequent use.

They are the result of observations, experiments, investigations, simulations, and calculations. Data is available in different types and kinds.

These are for example:

- Numbers (whole and natural)
- Boolean numbers (values 0 or 1)
- Characters and strings
- Signals (EKG, EEG)
- Images (camera, sono, x-ray, MRI, MRT)
- Sounds
- Number formats (calendar date, time, currency, weight, temperature, size)

As in all areas of life, data are of particular importance in healthcare. The associated unit and documentation / description are required in all cases. So the number 95 does not say anything on its own. In connection with a blood pressure measurement, the number has a meaning as, for example "Diastolic blood pressure measured in Millimeter mercury mmHG ". The application
eg the WHO criteria turns it into the information "elevated blood pressure".

Data, knowledge and information are interlinked. The starting point is the data, which must be precisely defined. This includes the origin, e.g. measurements by devices and people for treatment or studies, routine information systems or legally stipulated messages.

Data is recorded in order to be (further) processed. This is done with data processing systems that require data as input in a digital form with specified formats. The internal storage is often different and depends on the programs and operations used. External storage also depends on the application program, the storage medium and the destination of the storage. The exchange formats of data between applications, programs and facilities are more important. There are standardized procedures for this, which are based on different models.

Data type	Range of values	example
Boolean	True or false [0, 1]	Query about a gray value in an image (the gray value in the image exceeds the value 98 => true or false?)
Char	Single character from hexadecimal range [0000 to FFFF]; Special characters, numbers, letters, ... (depending on the *Coding)*	Example: Comparison of values with the characters "<" or ">" (is 98 < selected gray value range x?)
Integer	Integer values (32 bits)	Blood pressure values in mmHG are generally given as integer values
Real (float / double)	Floating point values (32/64 bits)	Example from signal analysis: amplitude (frequency) of 68.357 Hz

Knowledge

In health care a variety of knowledge based decisions are made when treating a patient. Education and training is the basis for imparting knowledge. Mainly textbook knowledge, empirical knowledge and up-to-date findings in scientific research are communicated.

In patient treatment, knowledge is used for diagnostics, therapy and prevention. The knowledge is stored in different formats. Digital formats are suitable for use in knowledge-based systems, e.g. expert systems or decision support systems. In conjunction with the patient's digital data, the systems can draw conclusions.

Types of knowledge

Medical (specialist) knowledge can be classified according to its type. Generation, storage, transmission and use play an important role in this.

Factual knowledge (Textbook knowledge) forms the basis of education and training. Examples in the field of medicine are body functions, anatomy, drugs and other forms of therapy. This knowledge is constantly updated and is available in textbooks, reference works and other printed materials, both in paper form and on digital media.

Empirical knowledge (external and personal experience) represents the second important basis of patient treatment.

Foreign experiences are often documented in the media and passed on in this way. Stone tablets and papyrus rolls are just as much a part of this as books and publications in printed form or on digital media.

Medical action is also based on personal experience, which is documented in notes or patient files. This documentation is necessary from a legal point of view, but also serves as an individual "reference work".

Special knowledge results from studies or data in patient records. Terms such as evidence-based medicine or case-based reasoning are examples.

Information

The interpretation of data - using knowledge - creates information. An information cycle consists of the steps of "observing the patient, collecting data, interpreting the data, drawing conclusions, taking action, observing again." In practice, the term is often not used in this definition. Data is referred to as information. New knowledge can also be developed from the interpreted data through generalization or induction.

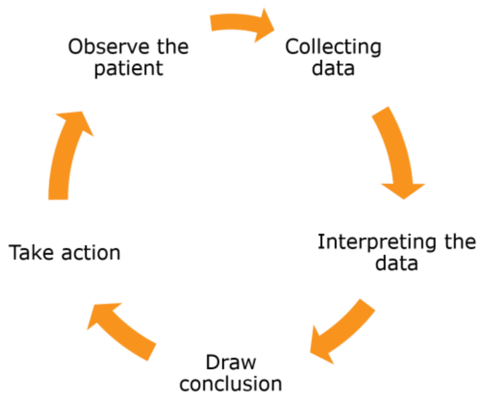

Figure 1: Information cycles[1]

[1] Source: Musen & Bemmel, 2002

Management of data, knowledge and information

Managing data is a process that has a long history in documentation. The recording of the data, the information and the conclusions in paper form allowed them to be saved for their own purposes and to be passed on. In the comparable form using a computer there are clear advantages:

- Security,
- Quality by checking the data input and
- electronic communication.

An important goal of the management of data, knowledge and information is their provision for the processes in the health care system, especially in diagnostics and therapy. For this purpose, specific methods, concepts and measures were developed.

Data structures and files

The storage and organization of data is defined in data structures. The simplest form is the file, which is stored digitally on different media. There are different functions of files, e.g.

- Storage of patient records,
- Reference files (index files, master patient index),
- System files for the operating system,
- Application files / applications.

Files can be open, protected, and encrypted. The simplest form consists of one or more records. A data record can consist of one data element or contain complex data elements, for example an electronic doctor's letter. The data element can be of a fixed length or of variable length.

The files are accessed using commands in sequential form, ie starting with the first data record and then all the others one after the other. Individual data records can be accessed directly via an index (index-sequential access). Another possibility is to use an algorithm to calculate the location of the data set in the file.

Database Management System (DBMS)

A medical information system often consists of a large number of data structures that require complex management. This is supported by a database and a database management system (DBMS, Database Management System). The database maps the logical structure of the data of an application on a physical memory, for example in files, with a definition language (Data Definition Language / Data Description Language, DDL).

A DBMS has, among other things, standard functions:

- Access protection, e.g. controlled by a language (Data Control Language)
- Competitive access, access by multiple applications / users
- Data integrity
- Data backup (backup and restore)
- Application-related logical views, which means that subsets of the data are provided
- Easy report generation, export of data
- Application program interfaces
- Data manipulation using a language, including a query language (Data Manipulation Language, DML)

Information systems

Medical information systems are an essential part of the health system. They support processes in the various areas of health care and the people and machine / technical resources involved in the processes. Information systems can be defined by their task and their technology.

Information systems technologies are for example:

- EHR based systems
- Mobile systems
- Web based systems
- Knowledge-based systems (decision support)
- Video consultation

[2] electronic health record

Information and communication technology

In keeping with technological developments, the information and communication technology (ICT) used in medicine has changed dramatically. While a few years ago the use of video consultation was still a research topic, such as the presentation of a patient with diabetic foot to a surgeon, nowadays there are complex systems, for instance in stroke care with high-resolution screens, which significantly enable the success of treatment in severe cases through a dialog between the patient, the local physician and a specialist in a stroke center.

In addition to the intrinsic quality of the devices used in terms of screen resolution, performance and operating features in information technology, it is above all the diversity of the applications available (a.k.a. apps) that make them suitable for widespread use.

The individual components of ICT technology are

- the user interface consists of a screen with input options via keyboard, mouse, touch screen, speech, external device such as sonography and output unit for text and images, speech and sounds
- the arithmetic unit with different processors for the actual arithmetic and the control of the screen and other units,
- the memory as main memory, image memory and data memory.
- the software as a control element for the various components and
- the communication techniques;
    - Physical transfer of data on data carriers, such as memory chips in the health card, memory sticks, external magnetic disks and CDs / DVDs, SSDs (Solid State Disc).
    - Forwarding (communication) in networks, either wireless (wireless LAN, cellular network) or wired.
    - Networks such as the German telematics infrastructure are a particular development. They offer specific access controls and other protective mechanisms and, through integrated functions, facilitate secure communication.

Interoperability

Interoperability is understood as the cooperation between devices and systems. The basic form is the technical interoperability of devices, such as external storage devices or medical devices. There are established standards for this, such as USB (Universal Serial Bus), which enable data transmission. In addition to the hardware, this also includes the programs that can operate these interfaces. They use a specified language (syntax), which is defined by format and content (control commands).

Semantic interoperability, which is characterized by different models and the meaning of the individual components, is important in medicine.

Interoperability enables the understanding of the data and is a prerequisite for meaningful further processing.

This also includes interoperability of the work processes that use the data. In its simplest form it is the representation of data, information and knowledge. Templates (stencils, templates) are used for this.

Interoperability is only possible when using standards that are defined between the parties involved. This can be done locally, regionally, in the facilities or through SDOs (Standard Development Organizations).

1.2 Purpose of collected health data

Health-relevant data are primarily "exchanged" between patient and doctor, being the typical doctor-patient relationship. However, large amounts of data are also collected in the context of statistical health reporting or research projects. One example may be briefly described: the "National Cohort" (NAKO). The data are used for a health study lasting at least 5 years, in which 200,000 volunteers are examined a second time after 10-20 years. By examining genetic factors, environmental conditions, social environment and lifestyle, conclusions can be drawn about the development of diseases. "Strategies for better prevention and treatment of the most important widespread diseases should be derived from the findings."

Society as a whole benefits from

- the achievement of macroeconomic effects,
- making public health governance more effective,
- increasing the distributive justice of care and
- increasing the effectiveness and efficiency of scientific research and training.

Regions, geographically considered as economic areas, benefit from

- the achievement of regional economic effects,
- a promotion of regional supply equality and
- support for regional health governance.

Patients benefit from

- an improvement in the eligibility conditions and
- an improvement in the quality of care.

Physicians in the doctor's offices benefit from

- an increase in the quality of medical work,
- the reduction of costs,
- an increase in the effectiveness and efficiency of the office organization,
- a strengthening of competitiveness and
- better use of education and training opportunities.

Hospitals benefit from

- an increase in the quality of care,
- lowering costs,
- an increase in organizational effectiveness and efficiency,
- strengthening competitiveness and
- strengthening the effectiveness of research, education and training.

Payers benefit from

- the reduction of expenses for utility services,
- the reduction of administrative costs,
- the effectiveness and efficiency increase of the internal cash organization,
- strengthening the participatory function in health policy governance and
- strengthening competitiveness.

It is therefore a broad mix of beneficiaries who profit from the collection and analysis of medically relevant data.

[3] https://www.gbe-bund.de/

[4] see. http://www.faz.net/aktuell/wissen/medizin/datenspeicherung-in-der-gesundheitsstudie-nationale-kohorte-13259109.html or. http://www.nationale-kohorte.de/

Possible uses of health data obtained

Health data is an important factor in the health system. They are often referred to as the "gold" of the health system, as research on diseases is only possible on the basis of a large amount of (health) data. From an economic point of view, this research is certainly for the benefit of companies - but also for the benefit of the patient.

The health data therefore have a certain added value for the "users" of this data. For example, health insurance companies can offer their services more "economically" only because of the broad access to health data.

Big data

According to LANGKAFEL, Big Data means "the acquisition, storage, search, distribution, statistical analysis and visualization of large amounts of data5." In the - predominantly IT / technology-heavy - specialist literature6 the term Big Data is also associated with the so-called 3V plus 2 brought.

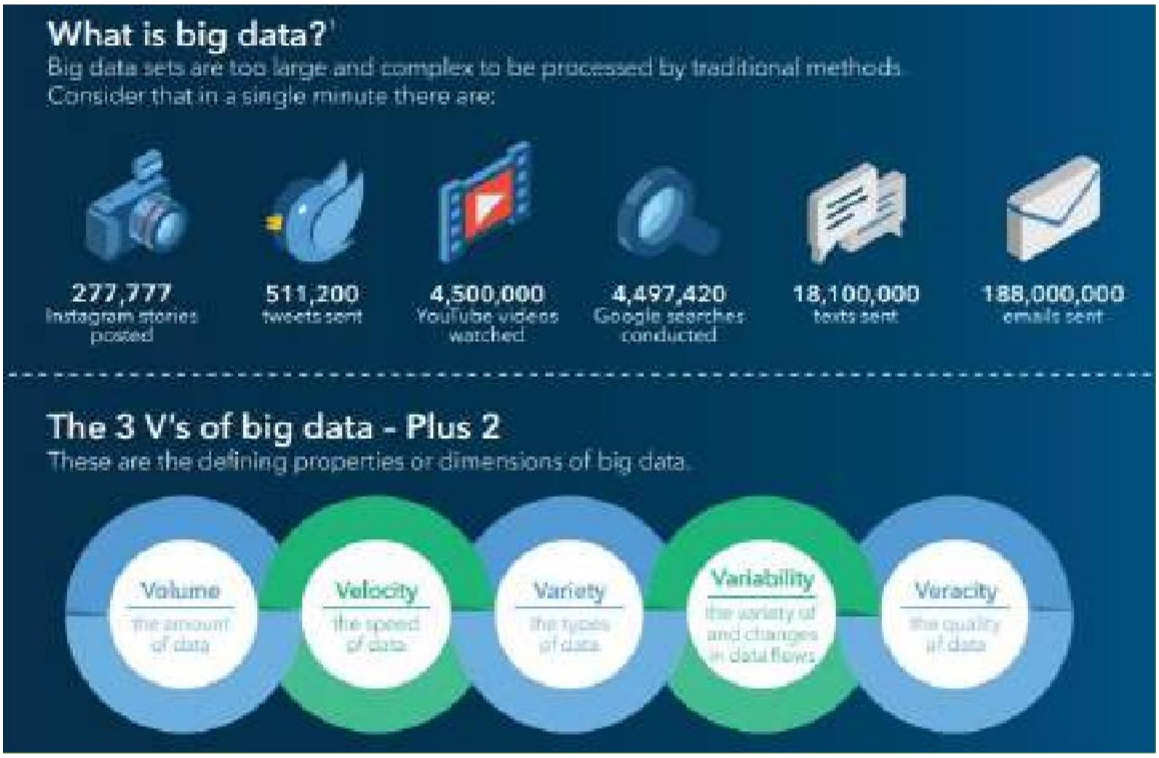

Figure 2: The three Vs plus 2 of Big Data[7]

[5] see LANGKAFEL, 2014
[6] see, for example, MARKL, 2012
[7] Source: based on https://www.ibmbigdatahub.com/infographic/four-vs-big-data

- *Volume* (Amount of data) - describes the huge amounts of data ranging from several terabytes to zettabytes. With such large amounts of storage, the data is usually stored in virtualized clusters or in the cloud. The (storage) servers are controlled within the IT infrastructure as required.
- *Variety* (Data diversity) - illustrates the heterogeneity in which the data is structured. The data mostly exist in different structural forms and formats and are available in different source systems.
- *Velocity* (Processing speed) - describes the speed at which the data is processed and analyzed. Processing can either take place in real time or near real time.
- *Veracity* (Quality) - Data with high veracity has many records that are valuable for analysis and that contribute in meaningful ways to the overall results. Low-truth data, on the other hand, contains a high percentage of meaningless data. The non-value in these data sets is referred to as noise. An example of a high-truth data set would be data from a medical experiment or trial.
- *Variability* (Variability) - describes the variety and change in the data streams. Companies have to cope with new trends in social media and peaks in the data load.

Data analysis is a great challenge because data more often than not are located on different systems and frequently use different data formats. Establishing a cross-connection between the data is therefore not easy to achieve. "The requirement for big data is to extract data from simple text files and to combine it with audio or video data." When analyzing data, accuracy is initially only secondary. It is much more important that the large amounts of data can be classified into trends, probabilities and development patterns. Prediction models and optimization algorithms then ensure that the large amount of information is processed quickly, so that final precision can also be achieved.

[8] One terabyte (TB) is equal to 1000 gigabytes (GB). A zettabyte is equal to seven times the exponential of 1000, which can also be written as 1021. For comparison: according to various estimates, the World Wide Web was about half a zettabyte in size in 2009; see.http://www.theguardian.com/business/2009/may/18/digital-content-expansion.

[9] Structured data (tables, entries), semi-structured data (e.g. XML), complex structured data (hierarchical sources; master data), unstructured data (voice, videos)

[10] see KOUT, 2013, p. 6

[11] Source: https: //www.sas.com/content/dam/SAS/documents/infographics/2019/en-big-data-110869.pdf accessed on March 1st, 2021

Health Care Analytics

The term health care analytics is virtually inseparable from big data, since it essentially refers to the types of assessments applied when using Big Data material. The term itself is borrowed as a product category from marketing and consulting services.

The promise of Big Data in the healthcare sector is to help contain the ever-increasing costs of medical care. The use and gains in knowledge from Big Data should lead to more effective medicine which is also of superior quality.

The key competence of Big Data analytics is the complete and correct collection of data as well as its interpretation and the accompanying decision support for physicians.

Big data, however, is not yet used everywhere - a special focus is placed on the clinics. In recent years, an informal level model for the adoption of big data in everyday clinical practice has therefore been established, the so-called Healthcare Analytics Adoption Model (HAAM).

As a general rule, the higher the level, the more likely a clinical facility is to implement the Big Data approach. Most facilities are already able to work very well and effectively at level 5 or 6, while those at level 3 or 4 are still working very inefficiently.

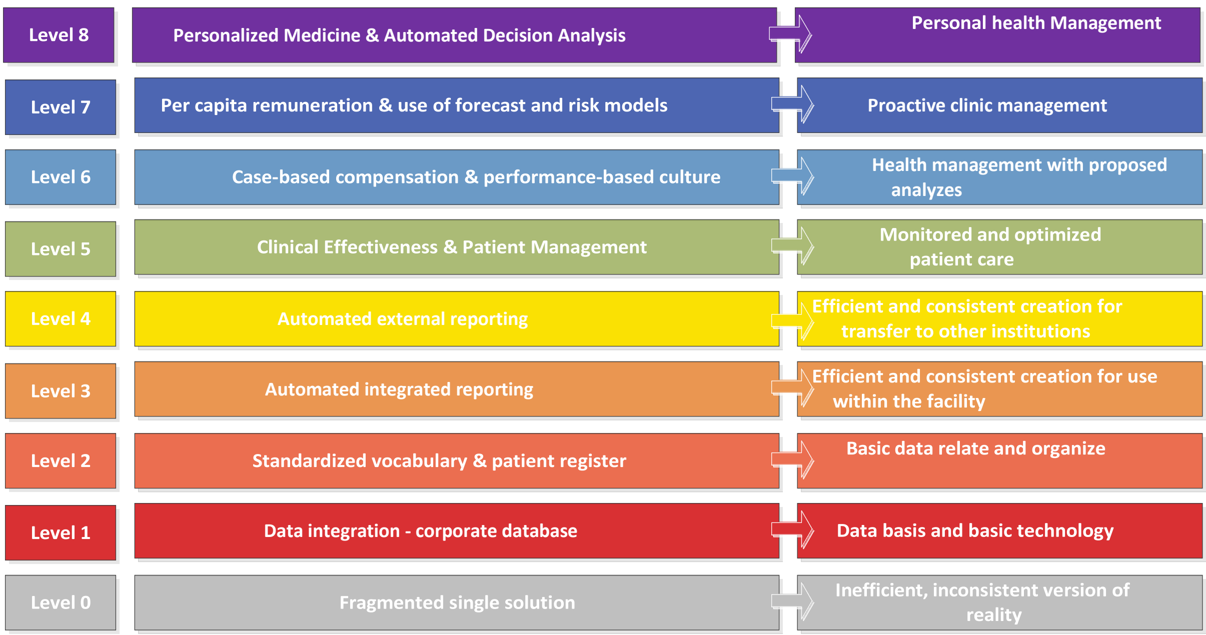

Figure 4: The eight levels of the Analytics Adoption Model[13]

[12] see pANDERS ET AL., 2013

[13] Source: RUSSOM, 2011, p. 6

The amount of digital medical data has increased dramatically within a very short time, which can be attributed in no small part to the electronic conversion of medical processes. A hospital stay results in a data volume that - according to LANGKAFEL - corresponds to around 12 million novels. In addition to the countless Internet discussion boards in which medical histories are exchanged and valuable information about the course of the disease and side effects lies dormant, databases already exist with millions of patients and the corresponding therapies.

But the highly complex processes in the human body can hardly be investigated using purely statistical methods and models.

“These are dimensions that without IT technology could no longer be managed at all. ”This is why big data (health care) analyzes sometimes fall back on prediction models (so-called predictive analytics).

Patients and caregivers should likewise be empowered by Big Data (in an indirect way) to practice preventive self-care. The end result is personalized medicine - medicine tailored to the individual patient. The use of big data is also advantageous in the area of decision-support systems. The more data that can be compiled from patient records, the more precisely a profound data and knowledge pool can be generated. The results from interlinked data combinations are currently of a pretty high quality.

Big data is being increasingly integrated into medical research. One example worth mentioning here is research in the field of genomics, which offers opportunities via the Big Data approach to make diseases that are incurable today potentially treatable in the future.

[14] LANGKAFEL, 2014. Imaging methods such as MRT or CT images should be mentioned, for example.
[15] From hospitals, insurance companies and private providers of genetic testing etc.
[16] see LANGKAFEL, 2014
[17] “indirectly” because a statistically secure basis for doctors and patients can only be created with a large amount of data (law of large numbers). The following principle can be noted: the more data is available, the more meaningful the results are.
[18] For this purpose, PIETSCH, 2013: “Big data offers a new approach to the investigation of complex phenomena. Because modern information technologies can process large amounts of data, statistical methods can be used that make predictions based on the original data. The detour via hypotheses that put a few parameters in a simple functional context would then be superfluous. An example of this so-called non-parametric approach are nearest neighbor algorithms that search for cases that are as closely related as possible in huge mountains of data, for example for a patient that is as similar as possible. It is about the vision of personalized medicine, in which each patient receives a treatment that is personally tailored to him, possible side effects can be individually predicted and better preventive care is possible, because everyone knows their risk factors. This contextualization, analogous to personalized internet search or online advertising, is a general feature of the new big data science. "; see alsoFigure 4 - Level 8.
[19].FELDMAN ET AL., 2012, p. 5
[20] http://www.spiegel.de/wissenschaft/medizin/big-data-wundermittel-auch-fuer-die-medizin-a-911333.html

Information security
Introduction to information security

With regard to information security when transmitting messages - such as the exchange of information between the physicians office (electronic hospital referral) and the clinic (electronic physicians letter) - there are central requirements in terms of data security and data protection. These are: [21]

- Confidentiality
- Authenticity
- Integrity

Explanation of terms
Confidentiality

In many cases, the content of communications is confidential.

It is noteworthy that many e-mail communications are transmitted without strong password protection or even unencrypted. Interception or eavesdropping on the communication or its content can be excluded if a secure transmission channel can be used and the communication is physically under the sole authority of the communication participants. In internal networks (intranet), this still appears to be uncritical and possible, since it is a closed system. However, this situation tends to get more challenging from a telemedical point of view, because it is precisely in this area that the objective is to establish cross-institutional communication. Special mechanisms must therefore be created to ensure the confidentiality of communication.

In principle, this can be done in two ways:

- A more secure communication channel is physically created in the form of a bilateral dedicated line that is not accessible to anyone else or
- the exchanged information is encrypted in a way that allows only the addressee to read them.

[21] see HAAS, 2006
[22] see HAAS, 2006, p. 51ff.

Authenticity

Another important aspect is the unequivocal identification of the sender and recipient of a message.

- Sender and receiver
  - If a doctor's office receives an electronic physicians letter from a hospital on the basis of which further treatment is carried out, the staff must be able to verify beyond doubt that this letter actually originates from this hospital or even from the medical professional indicated as the signatory.
- message
  - Received messages and the attached documents should be authentic. It must therefore be ensured that the stated identity of the author corresponds to the actual identity.
  - Procedures must be in place to enable the recipient to verify this.
  - The check of authenticity is not always relevant only at the time of the arrival of a message or attached documents, but it is also necessary that a renewed check must be possible at a later point in time.
  - If, for example, electronic documents are stored in a multisite electronic patient file, users who view this file at a later date - even years later - must be able to verify the authenticity of the documents stored in it.

Integrity

An additional aspect of electronic communication is the unforgeability and thus the integrity of messages. In addition to the authenticity of the sender, the recipient - in other words, a physician's office - must also be able to verify unequivocally that the message received as well as its attachments - for example, an electronic physician's letter - are still available in exactly the same form and with exactly the same content as they originally were when sent or signed by the originator

On the one hand, this requirement can be ensured by means of a digital signature; on the other hand, unauthorized modification during transport can be prevented by means of encryption procedures.

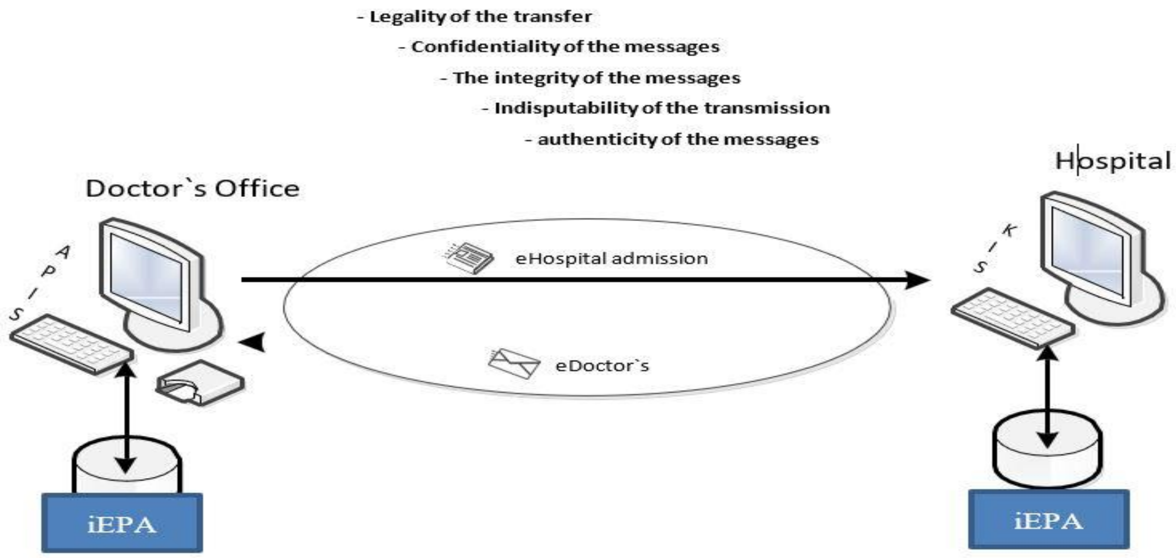

Figure 5: Messaging with legal aspects[23]

According to HAAS, "a computer network [...] can in principle be regarded as secure if it is not connected to an external infrastructure. Many hospitals therefore operate their hospital information systems in a closed network without a connection to any other external network or to the Internet. In some cases, special phone lines are used - sometimes with automatic callback by the communication partner - for example to transfer data to the health insurance companies in accordance with Section 301 SGB[24].
In principle, such a network is protected against foreign access - so-called intrusions - but only the elimination of all devices for the use of mobile data carriers such as CDs, floppy disks or USB sticks can prevent " malicious" programs from entering the network.
In this sense, the two most important categories of security risks are identified as:

- External intrusion into the computer network by unauthorized persons, who then pretend to be users of the network and spy out, copy or change data and
- Computer malware of various kinds, the objective of which is to obstruct operations, destroy the functionality of computer systems, alter or spy on data.

If a connection to other networks is therefore technically possible or necessary, it must be ensured beyond any doubt that all aspects of data protection and data security are still guaranteed. [25]

[23] Source: based on H.AAS, 2006, p. 50
[24] cf. http://www.gesetze-im-internet.de/sgb_5/_301.html
[25] H.AAS, 2006, p. 80

So if you want to establish trustworthy health telematics, a number of safety components are necessary. Data security can only be guaranteed when all components meet the security requirements26 and a comprehensive infrastructure has been set up. This includes how the following Figure 6 schematically shows the use of, for example, a single-stage protected computer network27. The computer network of servers and connected computers in a clinic, which is worth protecting (referred to here as an intranet), is subject to controlled access. The computers used within a demilitarized zone (DMZ) are sealed off from the intranet by the firewall. This separation allows on the one hand access to publicly available services (Internet, CMS, e-mail, ...), while on the other hand the computers of the intranet remain shielded from unauthorized access from the outside.

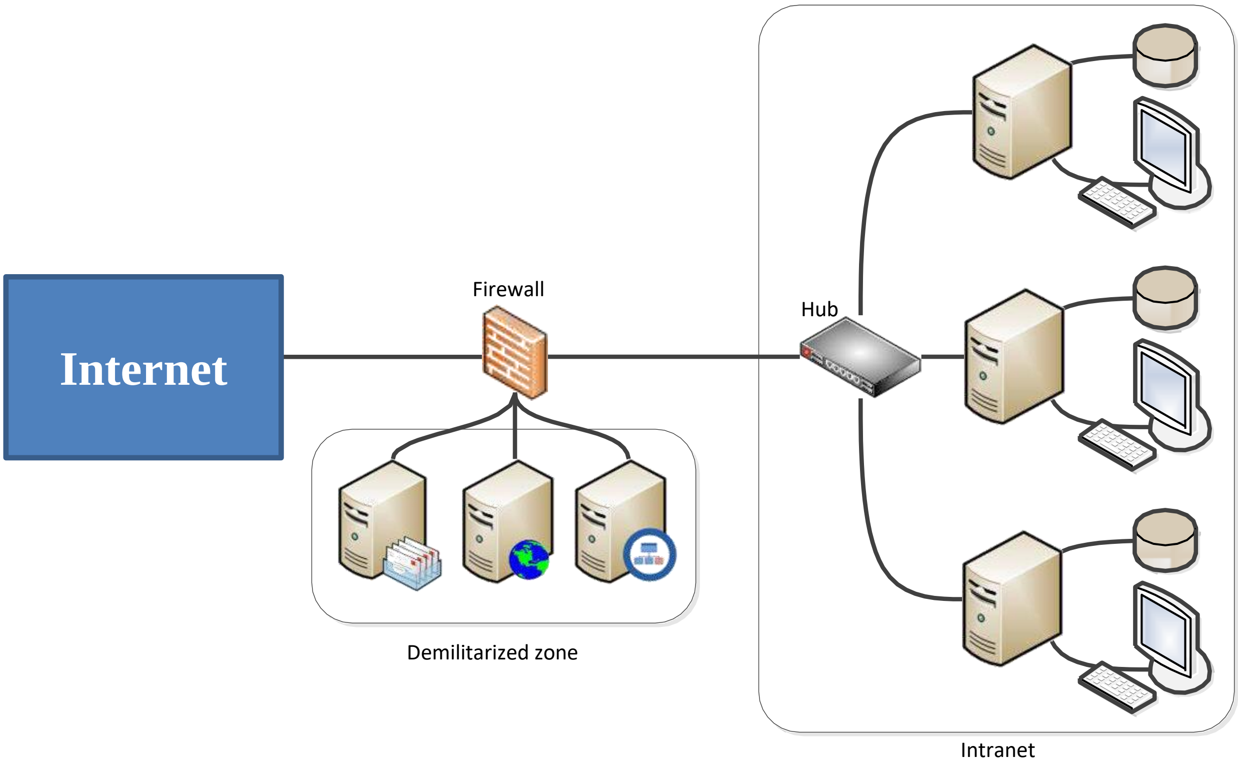

Figure 6: Exemplary use of a single-stage firewall with DMZ[28]

[26] see ISO standard 7498 (IT basic model) and ISO 10181 (security framework); see also requirements of the BSI https://www.bsi.bund.de/DE/Themen/ITGrundschutz/itgrundschutz_node.html

[27] There are also two-tier security systems. Here the intranet is again sealed off from the outside with an independent firewall.

[28] Source: based on HAAS, 2006, p. 50

1.3 Resilience of data

Data is used for statistics and decisions. The resilience is important for the respective purpose. This is a question of quality, which can be supported by various measures:

- Standardized data collection
- GCP (Good Clinical Practice) rules, such as drug trials
- Uniform work instructions SOP (Standard Operating Procedures)
- Quality criteria for information systems (issued by EuroRec or QREC)
- Data security also with corrections and deletions

Persistence

An important aspect of persistence is the ability to hold data (or logical connections) available over a long period of time. This takes place on appropriate non-volatile storage media such as memory chips (EEPROM in a “non-erasable” form and all non-transient chips29), hard drives, CDs, DVDs and even databases.

Databases are available in various forms. If the file system of the database is designed in such a way that the data can be stored persistently, the use of such a database is to be favored as large amounts of data can be stored in a secure manner. In this case, the protocol30 and the organization of the data storage are important: RAID systems (Redundant Array of Independent Disks) are preferred for this form of organization. If a system disk fails due to an incident, the redundant information is still backed up on a second disk. This data can be reconstructed and retrieved in the event of a data loss.

System breakdowns caused by errors or an unpredictable power failure during data transmission are problematic. Clear identification of the data is therefore all the more important. The unambiguous identification of the individual data records is therefore a prerequisite for the correct data to be stored in their information blocks.

[29] Such as the main memory (RAM). These buffered storage media lose their loaded information when the computer is switched off.

[30] The protocol should enable bidirectional and transaction-oriented transmission. In terms of data persistence, it is meant that individual data must be kept in a database management system (DBMS) until it is explicitly deleted. It must therefore be possible for the user to determine the lifespan of data, either directly or indirectly.

Reliability

Reliability refers to the dependability with which data is collected. It is a form of precision of a measurement and can be regarded as a quality criterion for the reliability of the data. Along with validity and objectivity, reliability is an important assessment parameter in relation to empirical studies or specific constructs. When measuring, reliability indicates the degree to which identical results are achieved under the same conditions. The higher this value, the more equivalent the results.

Reliability encompasses three aspects: [31]

- *stability* (Equality or similarity of the measurement results when used at different points in time)
- *consistency* (The extent to which all items that are combined into one characteristic in a test measure the same characteristic)
- *equivalence* (equivalency of measurements)

Validity

In the field of information technology, validity is a quality criterion that checks data for plausibility (sanity check). In principle, validation of data involves matching information between a desired state and an actual state. According to EUROSTAT, this refers to any activity with the objective of verification of the value of a data unit from a given set of permissible values. A measurement is to be regarded as valid if the collected values provide suitable key figures for the investigation. The validity can be classified according to different purposes or aspects of interpretation. [33]

[31] http://de.wikipedia.org/wiki/Reliabilit%C3%A4t
[32] http://ec.europa.eu/eurostat/de/data/data-validation
[33] http://www.wirtschaftslexikon24.com/d/validit%C3%A4t/validit%C3%A4t.htm

Literature and internet sources

Federal Office for Information Security: IT-Grundschutz, available at: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Kompendium/IT_Gr undschutz_Kompendium_Edition2021.pdf? blob = publicationFile & v = 6 (Accessed: 02/17/2021).

B.YOK , J. & A. CSAKI (Ed., 2013): Handbuch Digital Health. Practical guidelines for a networked health economy. Specialized publisher of the Handelsblatt GmbH publishing group: Düsseldorf.

ChannelPartner: Data Center, Modern Prediction Models and their Pitfalls, available at: http://www.channelpartner.de/a/moderne-vorhersagemodelle-und-ihre-tuecken, 3041596 (Accessed: 02/17/2021).

The federal health reporting information system, available at: https://www.gbe-bund.de/ (Accessed: 02/17/2021).

D.CORNER, O. (2005): Put everything on one card: Electronic government and the health card. In: Psychotherapists Journal 4/2005. Pp. 338-347.

D.ENZ, M. (2005): What do Open Standards have to do with eHealth, eGovernment and eGovernance? In: Swiss Medical Informatics 2005/55.

German Society for Medical Informatics, Biometry and Epidemiology eV - Medical Informatics Committee; https://www.gmds.de/de/aktivitaeten/medizinische-informatik/ (Accessed: 02/17/2021).

DocCheck Flexikon, definition of reliability, available at: http://flexikon.doccheck.com/de/Reliabilit%C3%A4t (Accessed: 02/17/2021).

E.NTERPRISE B.IG DATA F.RAMEWORK, AVAILABLE ONLINE AT: HTTPS: //WWW.BIGDATAFRAMEWORK.ORG/FOUR-VS-OF-BIG-DATA/ (Accessed: 11/24/2020)

EUROSTAT, European Statistics, Data Validation, available at: http://ec.europa.eu/eurostat/de/data/data-validation (Accessed: 02/17/2021).

F.ELDMAN, B., MARTIN, EM & T. SKOTNES (2012): Big Data in Healthcare - Hype and Hope, available at: http://de.scribd.com/doc/107279699/Big-Data-in-Healthcare-Hype-and- Hope (Accessed: 02/17/2021).

Frankfurter Allgemeine - Wissen: We do not collect data at random, available at: http://www.faz.net/aktuell/wissen/medizin/datenspeicherung-in-der- health-study-national-cohort-13259109.html (Accessed:02/17/2021).

Graduate College Knowledge Representation of the University of Leipzig, Decision Support Systems, available at: http: //www.informatik.uni-leipzig.de/~brewka/gk/teil6.html (Accessed: 02/17/2021).

HAAS, P. (2006): Health Telematics. Basics, applications, potential. Springer-Verlag: Berlin; Heidelberg.

KRAUS, H. (2013): Big Data - Fields of Application and challenges. Working papers of the FOM University, No. 41. MA Akademie Verlags- und Druck-Gesellschaft mbH: Essen.

KRCMAR, H. (2015): Introduction to Information Management. 2nd Edition. Springer / Gabler: Berlin; Heidelberg.

L.ANGKAFEL, P. (Ed., 2014): Big Data in Medicine and Healthcare. Diagnosis, therapy, side effects. medhochzwei Published by Heidelberg.

M.ARKL, V. (2012): Research, Innovation and Education on Big Data Analytics - Opportunities and Challenges. In: EBERPÄCHER, J. & O. WOHLMUTH (Ed., 2012): Big data becomes new knowledge. Münchener Kreis eV - supranational association for communication research.
Munich. Pp. 21-35.

National cohort - NaKo, Researching together for a healthier future, press release, available at: http://www.nationale-kohorte.de/ (Accessed: 02/17/2021).

Paul Ehrlich Institute (PEI) Central homepage of the institute, available at: https://www.pei.de/DE/home/home-node.html (Accessed: 02/17/2021).

P.IETSCH, W. (2013): Big Data in Medicine: Consultation hours at the supercomputer, available at http://www.spiegel.de/wissenschaft/medizin/big-data-wundermittel-auch-fuer-die-medizin- a-911333.html (Accessed: 02/17/2021).

Robert Koch Institute (RKI) Central homepage of the institute, available at: http://www.rki.de/DE/Home/homepage_node.html (Accessed: 02/17/2021).

Robert Koch Institute (RKI) Infection Protection - What Moves Us in Biological Hazards? , available at:http://www.rki.de/DE/Content/Infekt/Biosicherheit/Projekte/mobidig/mobidig_node.html (Accessed: 02/17/2021).

RUSSOM, P. (2011): Big Data Analytics. The Data Warehousing Institute Research - TDWI Best Practices Report, available at: https://tdwi.org/research/2011/09/best-practices- report-q4-big-data-analytics.aspx? tc = page0 (Accessed: 02/17/2021).

R.ÜPING, S. & N.GRAF (2013): Data Analysis: Big Data in Medicine. In: Deutsches Ärzteblatt 110 (41): A-1926.

SAS - big data. What it is and what you should know about it, available at:https://www.sas.com/de_de/insights/big-data/what-is-big-data.html (Accessed: 03/01/2021)

S.DIFFERENT, D., BURTON, DA & DJ P.ROTTI (2013): The Healthcare Analytics Adoption Model: A Framework and Roadmap. White paper. HealthCatalyst, available at: https://www.healthcatalyst.com/white-paper/healthcare-analytics-adoption-model/2/ (Accessed: 02/17/2021).

SCHILL, K.(1990): Medical Expert Systems. Oldenbourg: Published by München.

Spiegel Online Wissenschaft: Big Data in Medicine: Consultation hours at the supercomputer, available at: http://www.spiegel.de/wissenschaft/medizin/big-data- miracle-drugs-also-for-medicine-a-911333.html (Accessed: 02/17/2021).

The Guardian: Internet data heads for 500bn gigabytes, available at: http://www.theguardian.com/business/2009/may/18/digital-content-expansion (Accessed: 02/17/2021).

Topix Business Software AG, available online at: http://www.topix.de/ (As of: 02/17/2021).

Business dictionary, definition of the term validity, available at: http://www.wirtschaftslexikon24.com/d/validit%C3%A4t/validit%C3%A4t.htm (Accessed: 02/17/2021).