Wolfgang Giere Principles of an active Electronic Health Record (aEHR) Lessons Learned from BAIK

For my US-friends, the hardhats in admiration of VistA and their achievements

Principles of an active Electronic Health Record (aEHR) Lessons Learned from BAIK

Wolfgang Giere

Version 0.1 Jan 30, 2012

All rights reserved, ©Prof.em.Dr.med. Wolfgang Giere 2012

Dedicated gratefully to my wife, children and grandchildren especially to my staff, friends and teachers last but not least to the MUMPS community in shhort to all who supported me on the long way

Contents

List of Figures					
1	Abstract				
2	Introduction				
3	Prac 3.1	Practizing Physician and Programmer: Birth of a system3.1Reprogramming and Generalization			
4	The	DKD-system	9		
	4.1	Data Input and Structure: The DUSP System	9		
	4.2	Identification of a Form	9		
		4.2.1 Institution ID	9		
		4.2.2 Forms: Principle of Semantical Frames	10		
		4.2.3 Identification: Principle of Double Sequence	10		
		4.2.4 Episode-ID	11		
		4.2.5 Structure: Chapters within a Form	11		
	4.3	General Field Structure: Principle of Zipf's Law	11		
		4.3.1 Typed part: Ternary semantics	12		
	4.4	Data Types: Principle of Pragmatism	14		
	4.5	Text Generation System – DUTAP	14		
	4.6	Online Service for Physicians: DIPAS	15		
	4.7 Data Evaluation: The System IATROS		16		
	4.8	Minicomputers in Hospitals: DIADEM	18		
5	Mun	nps System: BAIK	19		
	5.1	The BAIK information model	19		
	5.2	Thesaurus	21		
	5.3	The difference between Collecting and ordering	23		
	5.4	IATROS	23		
	5.5	Adaptation of FileMan for BAIK	24		
	5.6	Graphical User Interfaces for BAIK	24		
	5.7	Achievements and Wishlist in the Eighties	25		

Contents

6	WorldWideWeb Services: BAIKweb							
	$6.1 \mathrm{Xmed} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots $	27						
	6.2 Dr. Antonius	28						
	$6.3 \mathrm{MedIAS} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots $	28						
7	VistA compared to BAIK	31						
8	Dream: WorldVista + BAIK features							
Bil	Bibliography							

List of Figures

2.1	Medical knowledge explosion until $1966[1]$	4
3.1	Newspaper, Jan 25, 1968: Author (left) and Horst Baumann, the pro- grammer, control the printed reports	7
$\begin{array}{c} 4.1 \\ 4.2 \end{array}$	DKD-Computer Center with the Siemens $4004/35$ Identification of a BAIK-Form: Principle of double sequence: forms per	10
	patient, form-versions per $group[1]$	11
$\begin{array}{c} 4.3 \\ 4.4 \end{array}$	Kodes for frequent, free text for rare findings – according to Zipf's Law[1] Illustration of Zipf's law counting the frequency of words in DKD-Diagnoses; HÄUFIGKEIT JEDES WORTES frequency of each word (of 429,665); ANZAHL VERSCHIEDENER WÖRTER count of different words (24,462); SUMMENKURVE summation: 0.75% of different words result in 52.8%	12
	of all words $[1]$	13
4.5	Interactive development with feed back and evolving complex target in- stead of a linear approach	15
4.6	IATROS Information Analyzing Text Retrieval Oriented Systemwith IATINT Query language for DUSP data[1]	16
4.7	Documentation Parameter $Set[1]$, for explanation see text on page 16	17
$5.1 \\ 5.2$	BAIK Information Model. Explanation in chapter 5.1 on page 19 Widmung of DIMDI honoring my role in the development of the ICD-10-	20
0.2	Thesaurus. (See footnote 5 in chapter 5.2 on page 21)	22
6.1 6.2	Search engine Dr. Antonius: No Results for <i>Wochendippel</i> Search engine Dr. Antonius: Result for <i>Wochendippel</i> with <i>Thesaurus</i> ticked showing 104 <i>Mumps</i> results, <i>Wochendippel</i> being synonym for <i>Mumps</i> 29	29

List of Figures

1 Abstract

Computer verändern die Medizin

(Manfred W. Gall, 1971 [2])

VistA is an outstanding electronic Health Record and patient management system, unparalleled and exemplary. It was developed in a decentralized, evolutionary effort together with or rather by the users (thanks to it's outstanding database and development environment FileMan). It is widely used, not only in the Veterans Administration where it originated.

BAIK is an (older) German electronic medical record system which has some unique features. BAIKweb is the most recent attempt to fulfil the goal of an active Electronic Health Record (aEHR). I is a prototype, but some of it's features could be desirable as future additions to VistA.

This publication is geared towards the VistA developers and responsible decision makers. It describes the result of more than forty years of development and modeling of medical documentation, information processing, classification, retrieval, selection, and presentation.

It contains many details and examples that were used in teaching. I will try to make the essentials very clear, to reevaluate them in the light of modern technology and the VistA example, and to repeat the resaults in concentrated form in the chapter 7 VistA compared to BAIK at the end. The chapters cover the development of BAIK from my first programming attempt 1966 until 2003, when I was emerited. Guided was my work by the models which are presented here, the Zipf Model and especially by my BAIK information model¹.

The very last chapter, 8 *Dream:* WorldVistA + BAIK features tries to outline further development of VistA using the lessons learned from BAIK and some features like e.g. Thesaurus-use for content-analysis, techniques to guide searches in the web and to filter the results to help the physician with appropriate support in the moment of decision etc. Thus VistA would become an *active Electronic Health Record (aEHR)*, giving the user much more² than he invested using the international community and leading edge technology.

¹This *Giere Model* has become canonical in Medical Informatics[3, pp 569 and 579].

²MuchMore was the working title of a transatlantic research cooperation financed by the US-National Science Foundation (NSF) and the European Union (EU) based on my BAIK information model[4]

1 Abstract

2 Introduction

Often I was asked why – as a physician – I started to work with computers. I cannot answer that question. But I know for sure that very early, already in 1957 as a first year student of medicine, I experienced the need for better information for the physician in the moment of decision. Had the doctor who maltreated me after a traffic accident used an active Electronic Health Record, I would not have spent many month (September to April) helpless in bed, lying flat on my back and missing a whole year of studies. The physician responsible for that disaster was - at least nominally - a specialist for both, surgery and orthopedics. And some years later I had discussions with my Doktorvater (supervisor of my MD-thesis) about the use of computers in medicine to overcome the rapid growth of medical knowwledge (see fig. 2.1) In 1968 he phoned me and asked whether I was still engaded in computer work. Then I had already brought into daily routine programmed reporting. So the answer was yes. And he told me of a position in Stuttgart at the Bosch-hospital for a physician who knew programming an IBM 360/30. I got it and switched to full time programming. In 1969 I got the responsability for the calculating center and organisation of the German Clinic for Diagnostic in Wiesbaden. From there I was called to the Goethe University in Frankfurt end of 1976. All the time I continued with my attempts to bring better information to the physician. I always knew that this required to know and understand what he documented in the medical record. Only this would enable the computer to know how to help the physician with his decision problem. The data of the patient has to be known and understood to search for relevant information for the actual decision. But how to get valid electronic data about the patient from the physician? The handwritten records were no help ...





Figure 2.1: Medical knowledge explosion until 1966[1]

3 Practizing Physician and Programmer: Birth of a system

Early in 1967 my new boss hired me see patients and to organize the new clinic for neuclear medicine¹ – it was the first one in Germany outside the universities – including documentation. When I presented him documentation sheets for punched cards – I have been very poud of them – he asked me what we needed (a key punch machine and a sorter) and who should do it: The physician had to fill in the forms and the secretary to keypunch them. Additional work for both, not to speak of the costs. He threw me out. The secretary saw my sad face and asked if it xould not be done otherwise: She had realized that we were dictating always the same phrases in the same order with only a few variations. She has had the idea of computerizing it. Without her, *Frau Wenz*, I would not have dared to engage on programmed reporting.

The typewriter with paper tape punch unit was borrowed. The IBM 360/30 at the city of Duisburg calculating center was interested. The system Programmer Horst Baumann helped me and taught me Assembly language. January 1, 1968 the system went into daily routine. And I had learned the first lesson: If you wish to get good data from the physician you have to relieve him of work, you have to reward him. The system must save time. It did: The doctor needed much less time, the secretary had to type much less and the reports were printed daily by the computer. Since there were no restrictions to add free text wherever needed, the programmed reporting was well received.

The system had positive side effects:

- The data were severely controlled for formal correctness and plausibility. Only if they were correct, the letter came out ready to be signed. If not a huge error message would show up in the middle of the text.
- The production of the letters immediately followed the end of the daily routine. The analysis of the test results and the nuklid scan was entered into the forms, it was typed and the paper punch tape was transported to the calculating center (via bicycle line connection). The next day when the patient came back to learn the results. he could take the printed report with him. I was checked and signed in his presence. (Compared to the normal waiting times for referral letters this was sensational)

¹Abteilung für Nuklearmedizin am Ev. Krankenhaus Bethesda in Duisburg, director Dr.med. H.A.E. Schmidt

3 Practizing Physician and Programmer: Birth of a system

- The content of the computer letters was well structured and allowed the receiving physician rapid orientation. Consequently they liked it inspite of the capital only script of the so called rapid line printers of that period.
- All data were stored electronically and could be analyzed.

The press was enthousiastic about the innovation 2 .

The program for this *Arztbrief* (referral letter) was huge, monolithic, written in IBM /360 Assembly language, several thousand commands, one punched card for each one³. However, it was well structured and had already subroutines for different data types. What is nowadays called *Natural Language Generation* (*NLG*) required to work with non fixed length variables throughout. That was unusual in those days.

3.1 Reprogramming and Generalization

When I later that year started to work in $Stuttgart^4$ I had to learn the Fortran language but continued to use an IBM /360 computer. My job was, to rewrite and enhance the MEDIUC automated diagnosis program for intoxications[5]. I did it successfully and learned a lot, including teamwork.

However, during those days I continued my electronic medical record work and generalized the principles of programmed reporting for both, data input and text output, DUSP⁵ and DUTAP⁶. Both became modular assembly programs, the basis for the DKD-system.

I redesigned the text generating and invented the language DUTAP. My resident Assembly kernel could interpret DUTAP programs fast and without prior linking! The DUTAP natural language generating language remained the same over the years⁷ It understood the ternary logic for branching, could interpret the different data types and was recursive using it's own stack.

²See picture 3.1 on page 7

³The compile and link run took a long time on the IBM 360/30. When we started the routine Jan. 1st 1968 there was still one error, a BNE (branch if not equal) instead of a BE (branch if equal). We corrected that error after loading the program by switching one bit at a certain address. Only after some weeks we reran the timeconsuming linking with a corrected statement.

⁴Medizinisch-Biologische Forschungsstelle am Robert Bosch Krankenhaus, Stuttgart, director Dr. R. Pirtkien

⁵DUSP is an acronym for *Datenerfassungs- Und Speicherungs-Programm*(data acquisition and storage program)

 $^{^{6}}$ DUTAP = Dekodierungs Und Text-Ausgabe Programm (Decoding and text output program)

⁷It was later implemented in Fortran and Mumps, see below 4.8 DIADEM and 5 BAIK

Bethesda und Rechenzentrum leisten Forschungsarbeit

Ein Elektronengehirn schreibt den Duisburger Ärzten Briefe

ph. Duisburg. Die teuerste Schreibmaschine der Welt wird zur Zeit vom Krankenhaus Bethesda benutzt: eine elektronische Datenverarbei-tungsanlage vom Typ IBM 360/30. Aber nicht Großmannsucht steht hinter diesem Beginnen, sondern der Versuch, in organisatorisches und medizinisches Neuland vorzudringen. Das Rechenzentrum der Stadt Duisburg unterstützt dieses von dem jungen Arzt Dr. Wolfgang Giere inspirierte Vorhahen. Inspirierte Vorhaben.

Vier Minuten braucht der Groß-computer der Stadt Duisburg, 70 mehrseitige Briefe an praktizierende Arzte zu schreiben. Er vergißt dabei nicht die höfliche Anrede "Sehr verehrter Herr Kollege" und "Sen verenter Herr Kolege" und die Schlußfloskel "Mit kollegialen Grüßen, Ihr sehr ergebener ..." Die Schreiben enthalten kompli-zierte Befunde und Therapievor-schläge für Patienten, die der nuklearmedizinischen Abteilung des Krankenhauses zu Spezialuntersu-Krankenhauses zu Spezialuntersu-chungen überwiesen worden sind. Diese "programmierten Briefe" er-sparen Krankenhausärzten rund die Hälfte, Sekretärinnen oder As-sistentinnen sogar dreiviertel der üblichen Schreib- und Diktatarbeit. Das Bethesda ist das erste Kran-kenhaus, das ein solches Verfahren anwendet anwendet.

Freie Hand gelassen

Freie Hand gelassen Eineinhalb Jahre hat 'Dr. Giere von der Idee bis zur Verwirkli-chung seiner Pläne gebraucht – und aufgeschlossene Helfer. Er brauchte einen Chef, der ihm freie Hand ließ. Er fand ihn im Leiter der nuklear-medizinischen Abteilung des Be-thesda-Krankenhauses, Dr. Schmidt. Und er brauchte eine hochergrüch-Und er brauchte eine hochgezüch-tete elektronische Datenverarbei-tungsanlage und Praktiker, die seine Vorstellungen in Computer-Programme umzusetzen verstan-den. Er fand sie im Rechenzentrum der Stadt Duisburg.

Nächtelang hat Dr. Giere mit dem 23jährigen Stadtinspektor Horst Baumann beraten und geplant, bis das Ziel erreicht war. Seitdem ist der junge Beamte mit einigen hundert medizinischen Fachausdrücken vertraut, und der Arzt wurde halb-wegs zum Elektroniker. "Ohne die-Unterstützung des Bethesda-Kran-kenhauses und des Rechenzentrums hätte ich mich an eine solche Aufgabe nie Dr. Giere nie heranwagen können", sagt

Mit Selbstkontrolle

Die mit der nuklearmedizinischen Abteilung des Bethesda-Kranken-hauses zusammenarbeitenden Ärzte haben sich inzwischen daran gehaben sich inzwischen daran ge-wöhnt, daß die meisten Briefe an sie von einem Computer geschrieben werden. Sie stoßen sich auch nicht daran, daß die Maschine nur über Großbuchstaben verfügt. Die Groß-schreibung ist praktisch der einzige "Fehler", den das Elektronengehirn macht. Es ist gewissermaßen eine ideale Sekretärin, denn es kontrol-liert sich selbst und kann sogar sinnwidrige Angaben in Schriftsätzen aufspüren. Trotzdem werden die Briefe von den Ärzten noch einmal durchgelesen, bevor sie abgeschickt werden. Auch der pertekten Technik gegenüber bleibt ein Rest von Skepsis.

Voraussetzung zur Verwirkli-Voraussetzung zur Verwirkli-chung programmierter Ärztebriefe war die Volldokumentation von, Krankengeschichten. Dr. Giere hat nicht auf die bereits von einigen Universitätskliniken erarbeiteten Methoden zurückgegriffen, sondern ein eigenes, erweitertes System entwickelt das wie er sart die ein eigenes, erweitertes System entwickelt, das, wie er sagt, die



Ein Arzt, Dr. Wolfgang Giere (links), und ein Beamter, Stadtinspektor Horst Baumann, leisteten Team-arbeit. Sie entwickelten den "pro-grammierten Ärztebrief". NRZ-Foto: Paetzold

Möglichkeiten eines Computers der dritten Generation besser nutzt. Das beinahe unbegrenzte " dächtnis" solcher Maschinen sy chert alle dem Arzt wesentlich "Gespeier. scheinenden Informationen über einen Patienten und erlaubt eine spätere wissenschaftliche Auswer-tung. In der modernen Medizin brauchen nicht mehr Tausende verstaubter Krankengeschichten verstaubter Krankengeschichten nachgelesen und mühsam ausge-wertet zu werden, wenn man stati-stischer Material der Forschung dienstbar machen will. Große Computer leisten diese Arbeit, die computer leisten diese Arbeit, die früher Monate und Jahre gekostet hat, in Minuten, — vorausgesetzt, daß man zur elektronischen Daten-erfassung übergeht.

Universelles Programm Dr. Giere hat sein Programm für ein medizinisches Teilgebiet, die Erkrankungen der Schilddrüse, zu-sammengestellt. Das von ihm ent-wickelte System ist jedoch univer-Schwierigkeiten auf andere Be-reiche ausgedehnt werden. Die Computer sind nicht zimperlich. Ihnen ist es gleichgültig, ob sie Befunde über Schilddrüsen- oder Gal-lenleiden speichern und zu einem Brief verarbeiten. Es kommt nur auf die Programmierung an.

Ein Computer-Programm setzt sich aus einer Vielzahl von Bau-steinen zusammen. Einzelinforma-tionen für diese Bausteine werden aus Fragebogen gewonnen, auf de-nen die gebräuchlichsten Diagnosen die gebräuchlichsten Diagnosen aus Fragebogen gewonnen, aut de-nen die gebräuchlichsten Diagnosen und Befunde, Therapievorschläge und Angaben des Patienten erfaßt sind. Der Arzt wählt unter den wie in einem Quizspiel vorgesehenen Möglichkeiten des Fragebogens aus und unterstreicht bestimmte Worta. Jedes dieser Worte enthält einen Code-Buchstaben, der zur Informa-tion für den Computer wird. Wenn eine Sekretärin bei der Auswertung des Fragebogens auf ihrer mit einem Lochstreifenstanzer gekop-pelten Schreibmaschine beispiels-weise die Buchstaben f-g-i-n tippt, macht der Computer für den Ärz-tebrief daraus den Satz "Bei der Untersuchung konnte man die feste, symmetrische, indolente Schilddril-se gut abgrenzen". se gut abgrenzen"

Erheblicher Effekt

Der Rationalisierungseffekt des neuen Verfahrens ist beträchtlich. Es wäre deshalb durchaus denkbar, daß sich in Zukunft Krankenhäuser zusammenschließen, um gemeinsam elektronische Datenverarbeitungs-anlagen zu nutzen, zumal sich die Computer nicht nur im medizini-schen, sondern auch im Verwaltungsbereich kostensparend verwenden lassen.

Figure 3.1: Newspaper, Jan 25, 1968: Author (left) and Horst Baumann, the program⁷ mer, control the printed reports

3 Practizing Physician and Programmer: Birth of a system

4 The DKD-system

1969 I started to work for the DKD¹ and transported my system to their Siemens 4004/35. I started with the two systems DUSP for data input and DUTAP for text output.

When starting the computer Center in the DKD^2 – it opened April 1970 – the physicians wished to use programmed reporting, too. We designed many data acquisiton forms and reports for different specialities from anesthesia to zytology. Due to the reduction of needed key strokes (codes for frequent findings and automated Ientification of the patient!) a secretary produced in one morning more than 300 pages of reports.

4.1 Data Input and Structure: The DUSP System

DUSP was an input generator controlled by *document parameter sets*. For each form these parameter sets described it's name, structure, the set of fields and their data types, formal and plausibility checks. I programmed and tested DUSP in its new form in Stuttgart, it was used in the DKD and later in DIPAS for many years. At the Goethe University Klinikum we reprogrammed it for BAIK using Mumps.

Let us look into the details which remained unchanged until today:

4.2 Identification of a Form

A form is the semantical unit. It's Identification requires the following IDs:

4.2.1 Institution ID

The reporting and documentation of medical facts can differ from institute to institute, from clinic to clinic, from school to school – as can the sets of Patients. Consequently the highest ID in BAIK is the GKZ^3 .

¹DKD acronym for *Deutsche Klinik für Diagnostik* (German Clinic for Diagnostic, called the *German Mayo Clinic*)

 $^{^{2}\}mathrm{DKD}$ = Deutsche Klinik für Diangnostik (German clinikc for Diagnostics)

 $^{{}^{3}}GKZ = Gruppenkennzeichen (Group ID)$

4 The DKD-system



Figure 4.1: DKD-Computer Center with the Siemens 4004/35

4.2.2 Forms: Principle of Semantical Frames

To be able to \langle understand \rangle automatically (electronically) the content of an entry we used semantical frames. Within a frame *EKG* a 78 in a specific field means pulse 78. in another context the same figure has a different meaning. Always the meaning is determined by the semantical frame and inside the frame by the position (field). The semantical frame in BAIK is determined by the so called AWZ^4 and the Version number.

4.2.3 Identification: Principle of Double Sequence

Whereas the AWZ for a semantical unit, e.g. EKG, remains the same, the version number changes according to the evolving structure of the data. The EKG report might first be dictated, then acquired using a filled in form, than a dataset from the Pipberger EKG analysis program, than the one from a more recent apparatus and so on. Always it is an EKG. But the Version Number changes with the evolution of medicine. Whether I can use certain data for research purposes has to be analyzed from case to case.

Consequently we get one sequence per semantical Entity, e.g. EKG

The other sequence, of course, is the number of EKG reports, a specific patient has, called LNR^5 .

It may well happen that a patient has reports with Version number 1, none with 2, several with three and so on. The two sequences are different. The one is bound to the biographie of the patient, the other to the development of medicine or reporting in a specific institution.

 $^{{}^{4}}AWZ = Auswahlzeichen (Retrieval-ID)$

 $^{{}^{5}}LNR = Laufende Nummer (running number)$



Figure 4.2: Identification of a BAIK-Form: Principle of double sequence: forms per patient, form-versions per group[1]

4.2.4 Episode-ID

Many forms can be attributed to one medical episode, a chemothapie e.g. or a specific study. this KVZ^6 is optional.

4.2.5 Structure: Chapters within a Form

Within a form (GKZ,AWZ,VNR) there may be one or more chapters. Each chapter allows for repetitive structures – very much like subfiles in FileMan.

4.3 General Field Structure: Principle of Zipf's Law

The smallest addressable unit of a semantical unit is called a field. In BAIK the general structure of a field is different from FileMan in order to comply with Zipf's Law: In every field after the typed or coded part a free text $Zusatz^7$ is allowed (see Abb. 4.3). It is separated from the first (coded or typed) part by an asterisc. The samantical structure of both parts is different:

 $^{^{6}\}mathrm{KVZ}=\mathrm{Konsulationsverzeichnis}$ (Consultation ID)

 $^{^{7}}$ Zusatz = second part, optional

4 The DKD-system



Figure 4.3: Kodes for frequent, free text for rare findings – according to Zipf's Law[1]

4.3.1 Typed part: Ternary semantics

The first, the typed part is similar to FileMan fields with a few differences: Every typed part has a ternary structure:

- 1. \gg = Nullstring: implicit negation
- 2. 0 = Zero: explicit negation
- 3. type = one of the many datatypes: positive content, to be decoded, if coded

The ternary logic is needed for the natural language generation and classification. Statistically it is not the same to have no answer, i.e. undefined, or an explicit zero (without pathological finding).

Zusatz: Binary logic

The second, the *Zusatz*-part, is binary: Either there is a free text or not. It is optional: There may be a text (of up to 99 lines) but it is not needed.

The Zusatz-capability allows the physician to mark any value. E.g. the height in cm might be given as 172*kyphoskoliosis. Than it depends on the type of a study whether the value will be included or not.



Figure 4.4: Illustration of Zipf's law counting the frequency of words in DKD-Diagnoses; HÄUFIGKEIT JEDES WORTES frequency of each word (of 429,665); AN-ZAHL VERSCHIEDENER WÖRTER count of different words (24,462); SUMMENKURVE summation: 0.75% of different words result in 52.8% of all words[1]

÷

4 The DKD-system

Another example of free text Zusatz is S^* after sphincterotomy. S stands for sphincter starr, in this case after an operative procedure. The binding of free text to a specific semantic frame allows short texts with well defined meanings.

It also allows to add personal remarks, unusual findings etc. In principle every field can have a *Zusatz*. It's meaning is defined by the semantical fram of the field.

4.4 Data Types: Principle of Pragmatism

The datatypes were chosen to minimize the needed key strokes at input time. All data types are well defined in Backus Naur formalism[6].

In this publication I only list the most frequent BAIK-datatypes. Details may be found in the literature[1]. Only interesting differences between BAIK and FileMan are pointed out.

- IKn Multiple choice, one answer per position
- **IKV** Multiple choice, multiple answers
- **IKZ** Time encoding (3w = 3 weeks, 12t = 12 Tage (days), 1j = 1 Jahr (year)
- **IKR** Radiomat Radiological dictation with code hierarchies, a very effective dictation tool
- **ITK** Thesaurus Control: For this field a thesaurus could be indicated, aganist which the words were matched. In case the word did not match, the typist had the choice of correcting the spelling or hit the terminology mailbox button. In that case the word was known in future. (The terminology mailbox war revised periodically)
- **INV** Numerical with Variable boundaries: Allowed the typist to override the realistic lower and upper limits if needed for unusual values. Think e.g. of an extreme liver weight with a big echinococcus cyst. Otherwise in medicine you would have to pull the window of allowed values wider and wider: Less and less plausibility control!

Like in FileMan so called document parameter sets contained the data types of every field including possible controls and boundaries.

4.5 Text Generation System – DUTAP

The output system was called DUTAP. Generation of quasi dictated reports from the kind of short hand entries – codes for frequent findings, free text for rare and atypical ones – was attractive for physicians and secretaries. It saved work for both of them. The idea was to reward the physician for good data. The computer program would proofread the data and print a big error message if there were implausibilities. Only if it was complete and ok, it would genmerate a nice looking report, ready for the physician to sign.



Figure 4.5: Interactive development with feed back and evolving complex target instead of a linear approach

4.6 Online Service for Physicians: DIPAS

This service for physicians was well accepted not only within the DKD but also by outside specialist physicians. Thanks to government grant money we could offer them online services in 1973. Twelve physicians with different specialties had online access to our DUSP/DUTAP system via the use of 300 baud lines and rapid online typewriters (ind of teletypes)⁸ Im Projektbericht[1] I described two main principles:

- 1. The principle of qualifying information preparation⁹ as model and longterm goal. (That is, what I call nowadays an active electronic medical record based on the information model.) The first step towards this longterm goal was the retrieval system IATROS (Information Analyzing Text Retrieval Oriented System).
- 2. Interactive development as introduction strategy.

Both goals were illustrated in the DIPAS-brochure[1] (see fig. 4.6 and 4.5)

Within this project the problem of classification and retrieval of the code/text-mix was solved:

⁸The project was named «Einführung der Datenverarbeitung in die ärtzliche Praxis – Dokumentation und Informationsverbesserung in der Praxis des niedergelassenen Arztes mittels EDV-Service (DIPAS)»

 $^{^{9}}$ qualifizier ende Informationsaufbereitung

4 The DKD-system



Figure 4.6: IATROS Information Analyzing Text Retrieval Oriented Systemwith IATINT Query language for DUSP data[1]

4.7 Data Evaluation: The System IATROS

The data controlled and stored by the system DUSP was well structured in forms (see 4.2 and Fields (see 4.3). Remember, as pointed out above, due to Zipf's Law the strored data consit principally of a mixture of coded or typed data and free text. To evaluate coded data and numericals is not difficulat. To evaluate text is more difficult, to analyse fields with a mix of coded and textual data requires special efforts. The original data have to be transformed into a metastructure that allows the combined evaluation. The transformation depends on the question to be solved. Every study has it's own needs. They may differ from each other. This is the reason for our strict distinction between primary patient data in the form/field-structure, as described above and the derived secondary data (see 5.1).

Within DIPAS all the principles were developed. We designed a methodology to build the secondary file for evaluation and an interpretive language IATINT to question it. This included the use of thesauri and semantical hierarchies for the searches.

The so called $\langle Documentations Parametersatz \rangle$ (Documentation Parameterset, see fig. 4.7) indicated how each field had to be transformed. It contained 24 bits¹⁰

• ENDANZEIGE Indicator of the last field to be evaluated

¹⁰Our computers in those days had max. 64K byte of memory and disks with 7 MB storage capacity. So we had to save bits!



Figure 4.7: Documentation Parameter Set[1], for explanation see text on page 16

- ZEILENNUMMER Line number
- FELDNUMMER Field number
- TYP DER KODIERUNG data typ of coding
- INTERPRETATIONSVORSCHRIFTEN Interpretation instructions
 - bit 17 Reserve
 - bit 18 ZUSATZ DOOKUMENTIEREN? With or without $Zusatz^{11}$
 - $-\,$ bit 19 STRUKTURTEIL NÖTIG? Is the identification of the semantical frame needed?
 - $-\,$ bit 20 ZEILENVERKNÜPFUNG? Valid for more than one Line
 - bit 21 FELDVERKNÜPFUNG? Valid for more than one Field
 - bit 22 INDEX FORTSCHALTEN BEI PUNKT? Augment index at period¹²
 - bit 23 INDEX FORTSCHALTEN BEI ZEILE? Augment index for new line¹³
 - bit 24 Reserve

¹¹for ZUSATZ see 4.3.1 on page 12, it is the uncoded free text addition allowed in every field.

¹²Within free text a special index counted the phrases. A special algorithm was used to find out the end of a phrase.

¹³used if arrays had the same semantics

- 4 The DKD-system
 - WAS = What?
 - WIE = How?
 - WO = Where?
 - WOMIT = Which Method?
 - WOZU = What for? (Purpose)

4.8 Minicomputers in Hospitals: DIADEM

The DUSP/DUTAP/IATROS system, heavily used in the DKD and as online service for outside physicians (DIPAS) was attractive for other hospitals, too. In a follow-up grant DIADEM ¹⁴ it was transported to minicomputers in different hospitals¹⁵.

Personally I am convinced if we were allowed to use Mumps (which we were not), DIA-DEM would have been more successful.

¹⁴Dokumentations- und Informationsverbesserung für den Arzt mit Dezentralem EDV-Modul (enhancement of documentation and information for the physician using decentralized EDP modules) (DIA-DEM)

¹⁵Interesting that we were not allowed to use Mumps, but were forced to use Fortran instead. We suffered greatly from the incompativility of the existing Fortran compilers. This Fortran version was later marketed as the first Doctors Office System in Germany. Neither system was successful.

5 Mumps System: BAIK

In 1976 I was called to the professoral chair for Documentation and Data Processing at the J.W.Goethe-Unviersity Klinikum in Frankfurt/Main. This gave me the freedom to redesign the system and reimplement it using Mumps. We did not change the data structure¹ according to Zipf's law, nor the semantical identification with the double sequence².

Our redesign was guided by the Information Model:

5.1 The BAIK information model

The BAIK information model is described extensively in all it's aspects in the BAIK-book[7] and in some detail in the BAIK-Chapter of Open Systems in Medicine[8]. For the convennience of the reader I quote the latter:

«

The Information Model – Cybernetics of Collecting and Ordering

The BAIK information model has guided the development for more than 20 years^3 . It still is the guideline for future development. ...

The backbone of the information model is a cybernetical information flow linking care, teaching and research together:

(1) A patient comes with a problem (?) to a physician. He examines the patient and notes symptoms, signs and tests in the medical record The medical record – next time – gives him the information he needs to treat (!) the patient. This is the classical, care oriented cycle.

(2) From the medical record data is selected to be classified and put into a register. The register in turn allows comparison of cases, adds comparative information of similar cases to the individual information of a single patient. This is the teaching oriented cycle, describing a multitude of similar cases. TEaching means to teach otheres or to gain insight oneself comparing similar cases.

(3) From the register statistical information can be drawn which allows the researcher to formulate a hypothesis. This in an experiment can be verified

 $^{^1 \}mathrm{see}$ 4.1 DKD system on page 9

 $^{^{2}}$ see 4.2.3 Identification on page 10 and figure 4.2 on page 11

³The quotetd text was published in 1995



Figure 5.1: BAIK Information Model. Explanation in chapter 5.1 on page 19

or falsified. The resulting knowledge again adds general information to the comparative and individual one.

Thus a complete cybernetical loop is formed between daily practice and scientific research. This "BAIK-Byke" allows to find the appropriate place for different constituents of a physicians workstation:

The acquisition of symptoms, signs and tests by the physician depends on his *experience*. The classification depends on a question which allows the correct establishments of differentiation *Criteria*. Tere is no general classification, classification alsways depends on a goal, never exists per se. Withour a specific question there is noch answer by classification. Classification means selection and appropriate transformation, always concentration and hence loss of details, gain on information about groups.

The statistical information of course is rendered by the interaction with *Methods*. this is the place for Systems like SAS, SPSS, BMDP and the like.

The general information is stored in electronic *Libraries*. The National Library of Medicine is a goog example. It renders additional information from the books.

The comparative information can be augmented by *Expert Systems*, computer aided instruction and rule based quality assurance. HELP^4 is a very good example for the data driven analysis which I had in mind when I designed the BAIK model.

The individual *Information Profile* is meant as a filter depending on the skills and interests of the physician, his previous knowledge an what was presented to him earlier.

What we wished to achieve was a mechanism by which the physician using BAIK would get – in addition to and triggered by the data of a patient he

⁴Warner, R, JD Morgan, TA Priyor, S Clark, W Miller: HELP – A Self-Improving System for Medical Decision Making. In: Anderson J, Forsyte JM (eds) MEDINFO, Stockholm, 1974

entered himself – automatically relevant information to his specific case. that could be a recently published article, help from a decision support system, information of other patients he had seen earlier and so on. The selection should take into account his speciality, interest, a specific selection of journals, in short: a physician-specific profile. We designed the system to use the world of electronically available information to help a physician. Again: A thesaurus was needed to translate medical record information into MeSH or MEDIUC or DXplain.

≫

5.2 Thesaurus

A thesaurus is needed for every cycle of the "BAIK-Byke" described above.

- It controls the input vocabulary at dictation or description. As a result every entered word is known to the thesaurus. If a string is new to the thesaurus the entering person is alarmed and has two choices: Correction of a misspelling or acceptance of a new word. In that case the word is entered in a *nomenclature letterbox* which is persiodically reviewed and voided. The terminologist adds the new word to the thesaurus with the appropriate semantical descriptions either as new synonym or as a completely new concept.
- It governs the classification. The information contained in the semantical frame (its context and location) has to be used to correctly bind items (numbers, codes, or words) to the semantical net, the thesaurus nodes.
- It is used for retreival. A physician may ask summary questions, e.g. "smoker", "viral infections" and gets answers regardless of the primary terms used in the relevant information. The thesaurus helps to interpret and translate the terminolgy of the question to the terminology of the entered patient data.

As mentioned already the thesaurus is also used as interface to the terminology of information systems like MEDLINE, rendering the appropriate MeSH-codes. The same is true for expert systems. Wherever different terminologies of systems have to be matched, the thesaurus is helpful.⁵

⁵Building of the thesaurus was started in the late 60s using punched paper tape. In the 70s we entered every dicated word of the referral letters of the German Clinc for Diagnostic in Wiesbaden. End of the 70s we installed BAIK in the J.W.Goethe-University-Klinikum for all dissection reports an other medical dictations[9]. By then the maintenance proved to be quite managable: one terminologist for half a day per month to work through the new entries in the terminology letter-box[10]. Our thesaurus was the basis of the ICD-10 diagnosis thesaurus which nowadays is maintained by the governmental agency DIMDI (the German analogue of NLM) and used in all German speaking countries (Germany, Austria, Switzerland) On the occasion of my 75th birthday DIMDI surprised me with a «Widmung» (see figure 5.2 on page 22

5 Mumps System: BAIK

Widmung

Habent sua fata libelli – Büchlein haben ihre Schicksale. Dieser Satz des Terentianus Maurus, vielleicht etwas aus dem Zusammenhang gerissen, gilt natürlich auch für das Alphabetische Verzeichnis zur ICD-10-GM.



Alles begann vor 15 Jahren in einem Projekt des Zentralinstituts für die kassenärztliche Versorgung. Unter der wissenschaftlichen Leitung von Prof. Dr. med. Wolfgang Giere, damals Leiter des Zentrums für medizinische Informatik der Universität zu Frankfurt/Main, entstand der ICD-10-Diagnosenthesaurus, ein schmales Bändchen mit vielleicht 10.000 Einträgen. Vorbild war Martin Luther: "Dem Volk auf's Maul schauend" sollte statt einer klassifikatorischen Kunstsprache die ärztliche Alltagssprache gesammelt und für die Kodierpraxis genutzt werden.

Über die Jahre hat der Diagnosenthesaurus sich gemausert zu einem umfangreichen Standardwerk mit über 100.000 Einträgen, das wohl jede Ärztin und jeder Arzt im deutschen Sprachraum kennt und – freiwillig oder unfreiwillig – immer wieder in die Hand nimmt.

Wir hatten das große Vergnügen, an diesen Arbeiten teilzunehmen. Wie kein anderer konnte Prof. Giere eine konstruktive Arbeitsatmosphäre schaffen, Probleme dank seiner langen Erfahrung pragmatisch aus dem Weg räumen und jederzeit sein Vorgehen theoretisch untermauern.

Für uns - in diesem Geschäft schon damals nicht ganz unerfahren - war das faszinierend, lehrreich, anregend und unterhaltsam.

Vor einigen Jahren hat das Zentralinstitut die Verantwortung für das Alphabetische Verzeichnis auf das DIMDI übertragen. Auch nach seiner Emeritierung stand uns Prof. Giere mit wissenschaftlichem Rat zur Seite.

Weshalb wir ihm die Ausgabe für das Jahr 2011 widmen? Nun, wer ihn gut kennt, wird um den Anlass wissen. Verraten wollen wir ihn hier nicht. Aber ein herzliches Dankeschön für die gemeinsame Zeit und die besten Wünsche für seine Zukunft sind uns wichtig. Ad multos annos!

Deutsches Institut für Medizinische Dokumentation und Information Köln

ICD-10-GM Version 2011

Figure 5.2: Widmung of DIMDI honoring my role in the development of the ICD-10-Thesaurus. (See footnote 5 in chapter 5.2 on page 21)

5.3 The difference between Collecting and ordering

Many more aspects of the BAIK-information model are explained in the BAIK monography [7], helas, in German only. But they are less relevant for the aim of this publication, for the lessons learned from BAIK for VistA.

This one, however, is essential in my opinion and I quote again the above cited publication of 1995[8]:

«Since we found that it is quite often overlooked, another aspect of the BAIK model shall be briefly mentioned: The difference between the data collection (medical record) and data order (register). The collection is unique, data is entered once. The form of the data may vary over time in spite of the fact that semantical frames in principle are rather stable. However, there can be and usually there are many registers per system depending on the ongoing research and appropriate classification criteria. A register is not just a "view" of the primary data. Selection and classification can imply complex transformations, including the normalization of different versions of data (see EKG example above⁶). A register contains metadata. The difference between data collection (medical record) and order (register) may be highlighted by these pairs of terms:

Data Collection (Medical Record)	Data Order (Register)
$\operatorname{patient}$	case
individual	group
characterizing	typing, classifying
$\operatorname{communicative}$	distributive
open for news terms	closed, predefined, standardized
lifelong	episode
primary	secondary

This list should provide a feeling for the fundamental differences of the two databases: The medical record with the primary cata and the register(s) with secondary, transformed, standardized and normalized data. They are by no means identical!»

5.4 IATROS

The Mumps system BAIK enhanced the functionality of the old IATROS system:

 $^{^{6}}$ on chapter 4.2.3 on page 10

- 5 Mumps System: BAIK
 - The classification used identical Document Parameter Sets (DPS)⁷, but allowed for Mumps code to control the transformations. Like FileMan it became a programming environment.
 - We constructed a powerful parser for the search commands. It replaced IATINT⁸. Especially the use of the thesaurus for semantical questions using the semantica net was greatly enhanced, including an interfaces to the widely used statistical package BIAS[11], developed by Ackermann in our Center of Medical Informatics.

5.5 Adaptation of FileMan for BAIK

All databases in BAIK used FileMan with one exception: The patient record file itself had a different design, was not transferred to the FileMan database. The main reasons for this were:

- different identifications (problem of double sequence)⁹
- different structure of a field (complex Field to comply with Zipf-law, ternary/binary, see chapter Field on page 11)
- different datatypes of fields¹⁰

We have had many discussions with hardhats and George Timson especially about means to reflect our different data structure to FileMan. It would have been feasable but clumsy. Consequently we did not change the well established format for the primary data, but secondary (derived, standardized and possibly transformed) data and all other files used FileMan. To reflect the double sequence identification structure – so important for semantical analysis – we invented for the thesaurus the *Kunstwort – artificial word* consisting of an identification praefix combined with the content of the field. These artificial words were used as synonyms to the appropriate terms in the semantical net.

5.6 Graphical User Interfaces for BAIK

There were some early attempts to adapt grafical user interfaces for BAIK, generators and others. Finally BAIK-web was the answer to that problem. Nowadays I would use EWD, of course.

 $^{^{7}}$ See figure 4.7 on page 17

⁸See chapter 4.7 on page 16

 $^{^{9}}$ See chapter 4.2.3 on page 10

¹⁰See chapter 4.4 on page 14

5.7 Achievements and Wishlist in the Eighties

In the eighties BAIK was used by several institutions and hospitals. In certain states (Bavaria, Bremen, Hessen, Niedersachsen, Rheinland-Pfalz) it was available free of charge. These states supportet the centralized maintenance (sufficient to finance one person). New development required additional capacities: Students who wrote their doctoral theses in medicine, computer science or medical informatics. At my institute I could promote physicians to *Dr. med.*, scientists to *Dr. rer. med.* and in cooperation with the Technical University Berlin Engineers to *Dr. Ing.*. This capability provided me with glqq slave labour». Step by step these doctorands worked on the wish list.

So far the requirements of model had been fulfilled only half. The automated feedback to the physician was not yet available. For instance the automated use of expert systems was not yet available, in spite of the fact that we designed successful expert systems already in the sixties and translated DXplain using the language translation software TRANSOFT[12], that Bill Moore invented for the translation of German medical texts into English. Together we developed the English to German version.

And in the meanwhile the world-wide web was invented and allowed for new functionality.

On the other hand software companies began to feed the growing market with glqq Hospital Information Systems». (And hospitals preferred to by software from a company than to get it from a university!)

At the J.Goethe-University Klinikm Frankfurt the routine operation of the calculating center was separated form the scientific institute, the *Zentrum der Medizinischen Informatik (ZInfo) – Center of Medical Informatics* freeing ressources for research and development. All of it was geared toward the fulfilment of the BAIK-model.

5 Mumps System: BAIK

6 WorldWideWeb Services: BAIKweb

In the nineties all open parts were identified and one by one solved by doctorands.

6.1 Xmed

The main disadvantage of IATROS was the fact that it answered only to direct questions. We were able to answer whether there were autopsies of HIV-patients before the illness had been named due to the semantical capabilities of the thesaurus-based search. But we did not get easily statistics of the morbidity of all patients. for this we had to classify the patient information and used ICD-9 and (later) ICD-10.

Automated classification of patient data containing much dictated free text (e.g. dissection reports) is a challenge. Of course the thesaurus, which knows every dictated word¹, is of great help. But you need to analyse rules. E.G. an aortic valve defect has an ICD-Code, a mitral valve defect another, but if both occur in the same patient at the same time, there is a third code for the combined vitium.

Xmed is a powerful system developed by two talented computer scientists[13][14]. Xmed has many capabilities:

- 1. It translates dictated German text into standardized German text. Standardized means
 - use of preferred terms only (thesaurus based)
 - isolation of medical facts (using conjunction-bound rules)
 - normalisation of the grammar of the isolated facts ((using grammar and POS rules)
- 2. analysis of the relations of medical facts to each other identifying
 - lokal relations (above, under, frontal, dorsal, ...)
 - time relations (same time, before, after)
 - causal relations (due to, causing)
 - intentional realtions (to care for, to prevent)

¹see above chapter 5.2 on page 21

6 WorldWideWeb Services: BAIKweb

(time, cause, intention)

- 3. encoding of facts into ICD (9 or 10) using facts, relations and ICD-rules
- 4. encoding of facts into German procedure codes (OPS) using OPS-rules

As one can see, in addition to the thesaurus many rule collections have to be into account:

- **Part-of-Speech (POS) rules** describing the structure of narrative German sentences and the resulting normalized sentence
- **Relation rules** describing trigger-words and the resulting relations. This quit frequently includes disambiguation of conjunctions with more than one meaning.
- **ICD-rules** for the choice of the correct code for any combination or certain circumstances (see example of vitia above)
- **OPS rules** for the choice of the correct OPS code.

To use the system in nowadays Germany, there would have to be added DRG rules as all reimbursement for hospitals is based on DRG. We designed a system to control whether all relevant facts to differentiate between possible DRG are documented and to remind the physician if decision criteria are missing[15].

6.2 Dr. Antonius

Dr. Antonius was a web search machine. Originally it contained a web robot searching German medical webpages $[16]^2$. They were analysed, descriptors put into a dadabase and searches enabled using the thesaurus. If you looked for *Wochendippel* you got only a few documents. But if you ticked the box using the thesaurus the search included the preferred term *Mumps* and the other synonyms and you got many hits.

The system was heavily used. A later version was designed to use google instead of our own web robot[17]. It was a fast system using GT.M and it's bitmap-capability. The idea behind the system was to provide the physician with actual information.

But ZInfo was dissolved³ and consequently the maintenace stopped. It could (and in my opinion should) be reanimated \dots

6.3 MedIAS

To help the decision making of a physician in view of a patient problem using computer capabilities has motivated me form the very beginning of my career⁴. The BAIK-model

²A major problem we had to solve to identify medical content was the neighbarhood to red light terminology.

³When I became emerited my position was not filled again with a medical informatics professor but with a clinician. The Center of Medical Informatics was closed. Some staffmembers changed to the calculation center.

⁴After I had become hospitalized for many month due to maltreatment, wrong decisions of colleagues.



Figure 6.1: Search engine Dr. Antonius: No Results for Wochendippel

🚰 Dr. Antonius - Die medizinische Suchmasc	hine - Microsoft Internet Explorer		_ 8 ×
Datei Bearbeiten Ansicht Favoriten Extr	as ?		1
⇔Zurück • → - 🔕 🖄 🚮 🥘 Suchen	🕼 Favoriten 🛞 Meden 🎯 🛃 - 🎒 🗹		
Adresse 🕘 http://www.dr-antonius.de		Hechseln zu	Links »
Dr. ANTONIUS Gesundheits-Server	Suchbegriffe (getrennt durch "OR", "AND" und 'NOT" und 'NEAR"): Wochendippel Thesaurus Suche Derzeit befinden sich 108917 Webseiten in der Datenbank.		•
 <u>Zur Suche</u> <u>Mveu - Version 2.0 - Neu</u> <u>Hüfe / Informationen</u> <u>Seite anmelden</u> <u>Kontakt zu den Entwicklern</u> <u>Über Dr. Antonius</u> <u>Haftungsausschluß</u> 	104 Treffer Seite 1 von 11 Ranking: 11.4 Ranking: 11.4 Adresse: http://de.dir.yahoo.com/Genundheit/Krankheiten und Beschwerden>Mumps Adresse: http://dx.dir.yahoo.com/Genundheit/Krankheiten und Beschwerden/Mumps/ Text: Revin Yahoo! S-Mail. Web-Site vorschlagen. Fläfe Yahoo! > Genundheit > Krankheiten und Beschwerden > Mumps Wallen Sie Ibr Ziel USA Australien Gemany-Mobil Griechenland Hong Kong Krostien Türkei Thaland Zairewe Ranking: 10.1 Tite: ArmPartner AG - KrankheitsLotze - Indexsuche A-D Adresse: http://www.artbachent.dd/.csi.phi/sang/krankheiten] app		
	Text: Ranking: 9.8 Titel: Krankheiten Adresse: HTTP //WWW NETDOKTOR DE/krankheiten/index asp?mode=alfa&letter=a Text: Ranking: 9.8 Titel: Krankheiten Adresse: HTTP //WWW NETDOKTOR DE/krankheiten/index asp?mode=alfa&letter=a		_
Britter CACHE	Lext Ranking: 9.8		

Figure 6.2: Search engine Dr. Antonius: Result for *Wochendippel* with *Thesaurus* ticked showing 104 *Mumps* results, *Wochendippel* being synonym for *Mumps*

6 WorldWideWeb Services: BAIKweb

 5 requires feedback to the physician. It had to be triggered by the classified patient data⁶ and would select appropriate information from the output of Dr. Antonius⁷, of expert systems and of similar cases.

To achieve this goal we first had to analyse and formalize the physicians information needs[18], than to construct a prototype using the actual web capabbilities[19]. The Medical Information Agent System (MedIAS) is the result. Of course, it uses the thesaurus, but also it uses profiles of the information needs of specific physicians. It knows whether he wishes to see original articles in New england Journal of Medicine or surveys and recommandations in the Deutsches Ärzteblatt olnly. It also counts how often the source had been presented to the physician already.

⁵See above chapter 5.1 on page 19

⁶Output of Xmed, see above chapter 6.1 on page 27

 $^{^7\}mathrm{See}$ previous chapter 6.2 on page 28

7 VistA compared to BAIK

First of all: VistA is successful and used by unnumbered people, maintained by the government and supported by *hardhats*, it is younger and up to date. BAIK is history, only a few installations still survive but will die due to lack of maintenance and active experts. The center of it's development, ZInfo, helas, is dissolved.

But some differences are of interest:

- **Double Sequence:** The sequence of changes in the structure of data (EKG interpretations e.g.) and the sequence of semantically identical reports delivered in different formats (EKG-interpretive-systems) reflected in the BAIK identification¹ in my opinion is needed for lifelong patient records. Only with this distinction of camparable data it can be dicided whether the data can be included in a study or must be excluded. VistA miraculously still is able to draw up the EKGs of Pipberger. EKGs ever since were stored in different formats. But as far as I know something like the principle of the *double sequence* is not available.
- Medical record and register distinction: The distinction between collecting and ordering, primary and secondary data² is strict in BAIK. For each scientific study an appropriate register can be generated keeping in mind that there is no correct classification for every use in medicine. Classification is a function of the question to be answered. And classification depends on time and sientific development. Who would have classified a stomch ulcus as infection 30 years ago. BAIK allows for this, VistA doesn't.
- Use of thesauri: VistA's data dictionary capabilities are great. BAIK was developed around thesauri³ and linguistic capabilities including automated translation (e.g. in $Xmed^4$). That is not the same, but the one does not exclude the other. Thesauri, Xmed and the like could enhance VistA. Especially the BAIK data-type ITK indicating thesaurus control⁵ for the input would be helpful.
- Augmented feed back to the physician: An active medical record delivers much more to the physician than what he himself has put into it. It analyses the patient data and searches automatically for relevant information which could help the specific physician keeping track of what he has seen already some times. This requires

¹See chapter 4.2.3 on page 10

 $^{^{2}}$ See above chapter 5.3 on page 23

 $^{^{3}}$ See chapter 5.2 on page 21

 $^{{}^{4}}$ See above chapter 6.1 on page 27

⁵See chapter 4.4 on page 14

7 VistA compared to BAIK

- $\bullet\,$ normalized search data (descriptors) as delivered by $\rm Xmed^6$
- \bullet a the saurus based web search technique like Dr. ${\rm Antonius}^7$
- an agent like MedIAS⁸ to keep track of the physician interest profile and it's use in the past, including an interface to decision support systems like DXplain[20].

 $^{^{6}}$ See above chapter 6.1 on page 27 7 See chapter 6.2 on page 28

⁸See chapter 6.2 on page 28 ⁸See chpater 6.3 on page 28

8 Dream: WorldVista + BAIK features

Already as a joung physician and pioneer of medical computer use I had a dream. As emerited professor of medical informatics I still have that dream: To provide the physician with actual, highly useful and well selected information for his actual decision problem to best help a patient. VistA is the best electronic medical record. with some of the BAIK features the dream could become true. Alls the clues are available.

It could be feasable – btw. it's all Mumps.

8 Dream: WorldVista + BAIK features

Bibliography

- Wolfgang Giere. Einführung der Datenverarbeitung in die ärztliche Praxis: Dokumentation und Informationsverbesserung in der Praxis des niedergelassenen Arztes mittels EDV-Service (DIPAS). gsf Gesellschaft für Strahlen- und Umweltforschung, Projekt Datenverarbeitung in der Medizin, DVM-Bericht 3, München, 1975. http://www.zinfo.de/DIPAS. ix, 4, 11, 12, 13, 14, 15, 16, 17
- [2] Manfred W. Gall. Computer verändern die Medizin. Fischer Taschenbuch, Frankfurt, 3. edition, 1971. 1
- [3] Thomas M. Lehmann and Erdmuthe Meyer zu Bexten, editors. Handbuch der Medizinischen Informatik. Hanser, München, Wien, 2002. 1
- [4] MuchMore. Web, 2001. http://muchmore.dfki.de. 1
- [5] Rudolf Pirtkien and Wolfgang Giere. Computereinsatz in der Medizin, Diagnostik mit Datenverarbeitung. Georg Thieme, Stuttgart, 1971.
- [6] Wolfgang Giere, Joseph P. Heger, and Norbert Krier. DOC-Programmierung. In Wolfgang Kirsten and Rüdiger Klar, editors, *Dokumentation und Informationsaufbereitung für den Arzt*, pages 371–380. epsilon, Darmstadt, Hochheim, 1996. 14
- [7] Wolfgang Giere. BAIK, Befunddokumentation und Arztbriefschreibung Im Krankenhaus. Media, Taunusstein, 1986. 19, 23
- [8] Wolfgang Giere. Medical Information Processing the BAIK Model. In E. Fleck, editor, Open Systems in Medicine, pages 24–34. IOS Press, Amsterdam, Oxford, Washington DC, 1995. 19, 23
- [9] P Röttger. Theoretische Grundlagen, empirische Generierung und Anwendungsstruktur eines Textverarbeitungssystems für die Pathologie. Habilitationsschrift, J.W.Goethe-Universität, Frankfurt, 1979. 21
- [10] Hubert Dettmer. Klartextverarbeitung von Sektionsberichten in der Praxis. Dissertation, J.W.Goethe-Universität, Frankfurt, 1979. 21
- [11] Hanns Ackermann. BIAS. epsilon, Darmstadt, sixth edition, 2010. includes programs. 24
- [12] G W Moore, U N Riede, R A Polacsek, R E Miller, and G M Hutchins. Automated Translation of German to English Medical Text. American Journal of Medicine, 81(1):103-111, 1986. 25

Bibliography

- [13] Christoph Luz. Xmed, vom Freitext zum Kode, Praxis der EDV-gestützten Klassifikation medizinscher Texte nach ICD und IKPM. epsilon, Darmstadt, 1997. 27
- [14] Andreas Gregori. Klartextverarbeitung als integraler Bestandteil medizinischer Dokumentation. epsilon, Darmstadt, 1995. 27
- [15] Björn Koschel. Entwicklung einer regelbasierten Validierung der Dokumentationsqualität zur Unterstützung der automatischen ICD-10 und OPS-301-Kodierung im Rahmen der G-DRG. Dissertation, J.W.Goethe-Universität, Frankfurt, 2007. 28
- [16] Wolfgang Kirsten, Michael Schlarmann, and Wolfgang Giere. Ein Prototyp eines deutschen medizinischen Webrobots als Teil des Frankfurter Index. In R Muche, G Büchele, D Harder, and W Gaus, editors, *Tagungsband, GMDS 42. Jahrestagung*, pages 316–320, Ulm, München, Sep 1997. MMV Medizin. 28
- [17] Winfried Bantel. Ein neuer Ansatz im Information-Retrieval Entwicklung, Implementierung und Test eines neues Algorithmus in der medizinischen Suchmaschine Dr. Antonius 3.0. Dissertation, J.W.Goethe-Universität, Frankfurt, 2003. 28
- [18] D Paparoditis. Ein Modell für Informationssysteme im medizinischen Umfeld. Dissertation, Technische Universität Berlin, 1996. 30
- [19] C Winkler. Profilgesteuerte automatische Informationsrecherche für den Arzt Konzeption und Implementierung eines Informationssuchagenten für heterogene medizinische Datenquellen. Dissertation, J.W.Goethe-Universität, Frankfurt, 2003. 30
- [20] G. Octo Barnett. DXplain. Web, Nov 2006. http://dxplain.org. 32