Health information systems in intellectual disability care: towards the re-use of routinely collected data.

Hubertus Johannes Martinus Tummers

Propositions

1. AI-based health care leads to better and cheaper care for people with intellectual disabilities. (this thesis)
2. There are more similarities between users than between systems. (this thesis)
3. An algorithm (a) is only as good as the data (d) it is trained on times the person (p) applying it. $q(a) = q(d)*q(p)$.
4. PhD candidacy an unemployment benefit for the highly educated.
5. Never postpone skiing.
6. The nitrogen crisis is a juridical crisis.

Propositions belonging to the thesis, entitled

Health information systems in intellectual disability care: towards the re-use of routinely collected data.

H.J.M. Tummers

Wageningen, December 14th, 2022

Health information systems in intellectual disability care: towards the re-use of routinely collected data.

Hubertus Johannes Martinus Tummers

Thesis committee

Promotors
Prof. Dr Bedir Tekinerdogan
Professor of the Information Technology Group
Wageningen University & Research

Prof. Dr Geraline Leusink, MD PhD MBA
Professor of Medicine ; Primary Health Care
Radboud University Medical Center, Nijmegen

Co-promotors
Dr Cagatay Catal
Assistant professor, Information Technology Group
Wageningen University & Research

Dr Hilde Tobi
Associate professor, Research Methodology
Wageningen University & Research

Other members
Prof. Dr S. Brinkkemper, Utrecht University

Prof. Dr C. Schuengel, VU Amsterdam

Dr K. Bevelander, Radboud University Medical Center, Nijmegen

Dr A. Fensel, Wageningen University & Research

This research was conducted under the auspices of Wageningen School of Social Sciences (WASS)

Health information systems in intellectual disability care: towards the re-use of routinely collected data.

Hubertus Johannes Martinus Tummers

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus, Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 14 December 2022
at 1:30 p.m. in the Omnia Auditorium.

Hubertus Johannes Martinus Tummers
Health information systems in intellectual disability care: towards the re-use of routinely collected data.
178 pages

PhD thesis, Wageningen University, Wageningen the Netherlands (2022)
With references, with summaries in English and Dutch.

ISBN: 978-94-6447-392-6
DOI: `https://doi.org/10.18174/576358`

It’s nice to be important, but it’s more important to be nice.

– Hans Peter Geerdes

Contents

List of Figures

List of Tables

General introduction | 1

1.1 Why this research?

Healthcare, software, these words not only rhyme, they often go hand in hand. Caregivers world-wide use electronic health information systems (HISs) to register their activities and for reimbursement of care provided. Yusof et al. [1] provide the following definition of an HIS: "HIS assist healthcare organizations to gather, process, and disseminate information within the organization and their environment. HIS incorporates various types of systems, including patient information systems, administrative systems, radiology and pharmacy information systems, telemedicine and hospital information systems, such as computerized physician entry systems." The adoption rates of electronic HISs compared to paper HISs mainly depends on two factors: developing countries vs developed countries and the type of care. For example, almost all (96%) general practitioners in Europe use an HIS in 2018 [2], while in sub-Saharan Africa very few HISs are used [3].

One type of care where many HISs are used is long-term care, although adoption of HISs in this type of care lags slightly behind other forms of care [4]. Long-term care includes care for the elderly, people with intellectual disabilities (ID), forensic care, youth care, and care for people with long-term illnesses. Long-term care requires a different approach for the HISs due to the long-term nature of this care (several years) compared to hospital and General Practitioner (GP) care (several weeks).

A special form of long-term care on which is focused in this thesis is the care for people with intellectual disabilities (ID). The American Psychiatric Association provides the following definition for an intellectual disability: "*Intellectual disability involves problems with general mental abilities that affect functioning in two areas: Intellectual functioning (such as learning, problem solving, judgement) and adaptive functioning (activities of daily life such as communication and independent living)*" [5]. According to the American Association on Intellectual and Developmental Disabilities

(AAIDD) the limitations must originate before the age of 22 [6]. ID affects about 1% of the worlds population [5, 7]. A less concise, although often used, definition of a person with an ID would be having an IQ lower than 70-75 [8, 9]. People with ID use a lot of care, and although their life expectancy is increasing, their life expectancy is still lower than that of the general population [10].

Data on persons with an ID is not available in just one system. The dossier is spread over a series of HISs because of the complexity of care for people with an ID involving an extensive set of care professionals. For example, they are in close contact with personal care aides who help them with their daily activities, but they also see a dentist or a general practitioner. [1] These systems also vary in functionality because of the wide range of care and cure people with an ID receive. For example, a person with an ID living in a care institution receives care from their personal care aide, who uses a system to register daily. But also, a general practitioner can be seen on the same day, who uses a different system to record visits. These HISs all contain and manage data about the patient/client, in the case of this thesis, the person with the ID. [2]

These registrations lead to a high registration pressure on care providers, for example, care providers in Dutch long-term care spend an average of 35% [11] of their time on administrative tasks, Swedish primary care providers 41% [12], and U.S. ambulatory care physicians 49% [13], a large part of which is in their HISs. These registrations are needed to ensure the continuity of care and that proper medical care is given. The re-use of this routinely collected data is limited. If the data is re-used, it is often limited to the use of data from one or two HISs or only on categorical or numerical fields, while a large share of data in (health)care is captured in unstructured textual reports [14, 15]. However, re-using routinely collected data in ID care from multiple HISs has great potential for scientific research, which can lead to improvement of care. Studies that use routinely collected data that could lead to changes in long-term care include fall prevention in home care [16–18] and dementia in elderly care [19]. In addition, this data has great potential for researching health over time, and for natural history studies amongst this vulnerable population [20, 21]. Re-using routinely collected in research can help

[1] Throughout this chapter, examples are shown using a fictional person Cherylbecca: Cherylbecca is 32 years old, has the Prader Willi syndrome and lives in care facility Care4ID for people with intellectual disabilities in the south of the Netherlands. Cherylbecca has a mild intellectual disability and is also in a wheelchair.

[2] Recently Cherylbecca does not want to eat anymore and is also very grumpy. This behavior is written down by her personal care aide in the electronic client file of Care4ID: myCaress. Her personal care aide does not trust the situation and makes an appointment with the intellectual disability physician for her. The ID physician examines Cherylbecca and suspects appendicitis, which she registers in her electronic patient file: OmniHIS. Cherylbecca is referred to the hospital where she undergoes surgery. This procedure is recorded in the hospital information system: HIX. Fortunately, Cherylbecca is doing well and is allowed to return to the care organization Care4ID. Here she goes through a rehabilitation process, which is logged in the electronic client file: myCaress. In the same month, Cherylbecca also went to the dentist and the physical therapist. These visits were registered in Dentalrules and SpotOnMedics.

close the gap left by the under-representation of people with disabilities in medical studies [22].

We describe the landscape of HISs in ID care with the help of the four Vs of big data: Volume (amount of data), Velocity (rate at which the amount of data increases), Variety (types of data and sources), and Veracity (reliability of data) [23]. To start with Volume: About 91,000 adults with an ID live in institutionalized care in the Netherlands [24], with often multiple registrations per day, the data increases with a tremendous Velocity. These people live in care facilities for multiple years, and therefore there is data from multiple years in the HISs, leading to an increasingly large dataset. [3]

To appreciate the Variety of data, one should realize that data is not only gathered in different systems, using different formats and standards, it is also of many different types. Some data is structured in categories, such as the International Classification of Diseases (ICD-10) [25], International Clasification of Nursing Practice (ICNP) [26] or Anatomical Therapeutic Chemical (ATC) classification [27] codes, but other data can be unstructured or semi-structured texts. Data in these HISs can even be images, sounds, or videos, showing the Variety of the data gathered [4]. The Veracity of the data indicates how reliable the gathered data is. Many different people work in the care for people with an ID, and there are just as many ways of how the data can be entered into the systems. Some care professionals may see the registrations as an obligation and shortly write something about the provided care. Other care professionals can be more extensive in their reporting and write down many details. Another Veracity issue is the use of "locally common" abbreviations, or even private/personal abbreviations, so abbreviations can vary from location to location and from person to person [5]. Human beings enter the data and make mistakes; sometimes data might be entered for the wrong client, or disorders or diseases might be misdiagnosed. All these factors influence the Veracity of the data.

Data as collected in ID care, has all properties of big data (4Vs). Many of the widely acknowledged problems with big data involve privacy and security. Privacy and security concerns limit the re-use of this routinely collected data. People with an ID cannot often decide about re-using their data themselves but rely on their legal representatives.

[3] **Volume:** Cherylbecca already lives in Care4ID for 10 years. All of her history is available in myCaress.
Velocity: Everyday there are at least two textual registrations for Cherylbecca, one in the morning and one in the evening.

[4] **Variety:** Cherylbecca has been very aggressive lately. Her personal care aide made a video of this behavior for the behavior therapist to assess.

[5] **Veracity:** Cherylbecca unfortunately has to move to another residential group within care organisation Care4ID. Her new personal care aide will be Bianca. Bianca reads the messages in Care4ID about Cherylbecca. She sees that the abbreviation "c/o" is often used. She does not know what this abbreviation means and calls Cherylbecca's previous personal care aide Monique. Monique explains that "c/o" means: care of. Now it is a lot easier for Bianca to read the files.

Also, the content of data from people with an ID can be very privacy sensitive by its nature [6]. A lot about their daily lives is registered, including the very personal parts. Because of the vulnerability of people with an ID, who are already experiencing more victimization and are at risk for social vulnerability [28], privacy and security concerns for a big data system must be guaranteed by its design.

[6] 6: Cherylbecca had a boyfriend living with her at the housing group. Unfortunately, her (not so nice) boyfriend broke up with her and Cherylbecca is upset about this. Her personal care aide Bianca has put the break-up in the electronic client file myCaress, so that the night care is aware of this.

People with ID often have difficulties expressing internal emotions and feelings [29], and telling a doctor if they experience pain [30]. Therefore, people with an ID often cannot explain their need for (health) care. With the help of data analysis techniques, some disease prevention, early detection, or maybe even predictions on the particular behavior of people with ID could possibly be done if data from multiple HISs can be combined. These analyses could help improve the quality of life for people with an ID. A lot of data is stored in the HISs used by the different practitioners, but almost nothing is done with this data except for reimbursement purposes. The potential of analyzing the large amount of data collected in ID care calls for the development of a big data system such that data from multiple HISs can be combined for research on this vulnerable population.

The empirical work of this project is set in the Dutch organization of ID care. This demarcation is done because of the unique organization of ID care in the Netherlands, where the ID physician plays a significant role. Furthermore, this is the country where the thesis was conducted and where the research team could efficiently work together with care institutions because of the academic collaborative "Stronger on your own feet." * People with an ID in the Netherlands often live in special care facilities, where there is taken care of them. Living in these facilities, people with an ID represent a large share of the healthcare use in the Netherlands [31].

---

* A collaboration between the Radboud university medical centre and sevenlager care provider organizations for people with intellectual disability `https://www.sterkeropeigenbenen.nl/english`

Concluding, it is important to explore how re-using routinely collected data in ID care can be facilitated. Because of the characteristics of these routinely collected data (4Vs), this thesis will focus on the prerequisites of a big data system. Therefor, the main objective of this thesis is to develop a framework for a big data system that combines data from multiple HISs and facilitates secondary analyses.

1.2 Thesis outline

To reach the objective, a range of studies that together form this thesis, was done. The chapters that report on these studies, their relation to one another and the thesis overall objective are depicted in Figure 1.1. This thesis outline is presented using the business process modeling notation (BPMN method). The BPMN method provides a graphical notation for specifying the business processes in a diagram [32]. The workflow is graphically presented in this figure, starting on the left side. The numbers in Figure 1.1 correspond to the separate chapter numbers, which should together lead to fulfilling the objective. The rationale for each paper is explained below.

Before data from multiple HISs can be combined for research, the healthcare landscape of ID care must be explored. Such an overview is needed such that it is known how the systems are structured, what data is being gathered, and how this is stored. This is called the architecture of the system. Furthermore, the challenges mentioned regarding the big data four Vs must be overcome.

Before data can be combined, it is needed to know what data is actually there. The objective of this study was to identify the key care professionals' experiences with, and desired requirements for, HISs. This identification was done in Study 1 utilizing a survey study. In this survey study, the data available in Dutch ID care and its peculiarities were identified. Using semi-structured interviews, a questionnaire was developed for care professionals working in ID care. Therefore, the state of the practice of HISs used in Dutch ID care can be found in Chapter 2.

In parallel with study 1, the state of the art of HISs world-wide is needed to make sure the provided framework is not only applicable to the Dutch organization of care (Figure 1.1). The objective of this second study was to assess the current state of HISs by means of a systematic literature review (SLR). By state the following is meant: mapping who is using the systems (i.e., stakeholders), where they are using it (i.e., domain), how they are using it (i.e., deployment), what they are using it for (i.e., features), and which problems are faced in this use (i.e., obstacles). This study is presented in Chapter 3.

The objective of study 3 was to develop a reference architecture for HISs following well-established architecture design methods. Each HIS is unique with its own way of structuring both its functionality and its underlying dataset. Due to this uniqueness, it is necessary to map the commonalities and variabilities between these HISs, such that it is known which parts are common between multiple HISs and can be re-used for research. Using results from study 1 and 2: a reference architecture for HISs was designed based on the views and beyond approach from Bass et al. [33]. This architecture can be used to map the application architecture of single systems. This design study can be found in Chapter 4.

Data does not combine itself (unfortunately). Since the data in ID care is very heterogeneous in formats, privacy-sensitive by nature, and large in size, it must be well thought out how to combine the data. The objective of this study was to map the to-be-built architecture for a big data analytics system. The reference architecture developed in study 3 was used to map a wide variety of HISs in different stages of development and acceptance used in Dutch ID care. An architectural design for a big data analytics system based on the similarities among these systems was developed. Furthermore, the challenges this big data analytics system faces are thought out and possible solutions are provided. This study is presented in Chapter 5.

Chapter 6 shows how, just like collected in intellectual Disability care HISs,large sets of textual data may be used for analyses. With a text mining approach, the possibility of analyzing large sets of texts is presented in Chapter 6. As a quick response to the COVID pandemic in the spring of 2020, the aim was to find clusters of articles that help researchers identify relevant research topics on coronaviruses and people with an ID. This clustering was done with the help of a text mining approach where Term Frequency Inversed Document Frequency (TFIDF) was used for the clustering.

In short: chapter 2 presents a survey study on HISs in the Netherlands, which, together with a Systematic Literature Review (SLR) in chapter 3, leads to a set of stakeholders, concerns, features, and obstacles. The outputs of these two studies are used as input for the study in chapter 4 in which a reference software architecture for HISs is developed. Based on this reference architecture, a big data system is designed for combining data from multiple HISs in chapter 5. In parallel we show the potential of big data analytics on textual data in chapter 6.

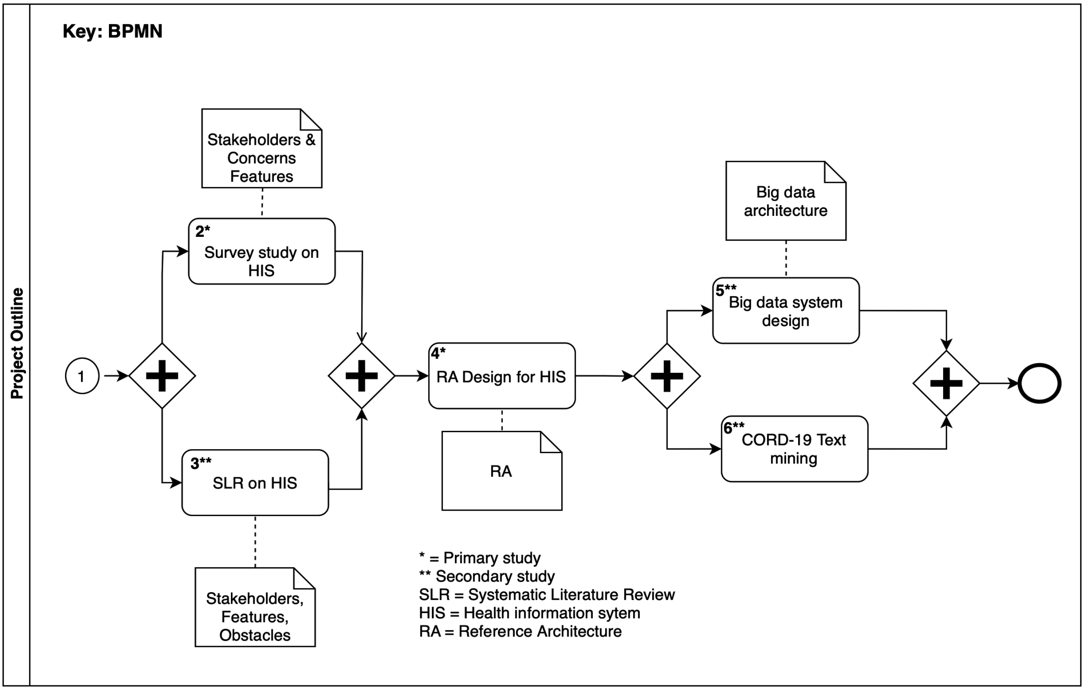


Figure 1.1: Thesis outline. Numbers correspond to chapter numbers. The start and end events represent the introduction and the discussion, respectively.

Landscape exploration

Study 1: Survey study on health information systems

2

Background: Care for people with an Intellectual Disability (ID) is complex: multiple health care professionals are involved and use different Health Information Systems (HISs) to store medical and daily care information on the same individuals. The objective of this study is to identify the HISs needs of professionals in ID care by addressing the obstacles and challenges they meet in their current HISs.
Methods: We distributed an online questionnaire amongst Dutch ID care professionals via different professional associations and care providers. 328 respondents answered questions on their HISs. An inventory was made of HIS usage purposes, problems, satisfaction and desired features, with and without stratification on type of HIS and care professional.
Results: Typical in ID care, two types of HISs are being used that differ with respect to their features and users: Electronic Client Dossiers (ECDs) and Electronic Patient Dossiers (EPDs). In total, the respondents mentioned 52 unique HISs. Groups of care professionals differed in their satisfaction with ECDs only. Both HIS types present users with difficulties related to the specifics of care for people with an ID. Particularly the much needed communication between the many unique HISs was reported a major issue which implies major issues with inter-operability. Other problems seem design-related as well.
Conclusion: This study can be used to improve current HISs and design new HISs that take ID care professionals requirements into account.

This chapter is based on a paper published in an international journal:

Tummers, J., Tobi, H., Schalk, B., Tekinerdogan, B., & Leusink, G. (2021). State of the practice of health information systems: a survey study amongst health care professionals in intellectual disability care. BMC health services research, 21(1), 1-9.

2.1 Background

Around 0.85 % of the Dutch population has an intellectual disability (ID) according to the definition of having an IQ below 70 [7]. People with an ID often live in long-term care facilities and have a large share in health care use in the Netherlands [31]. Care for people with an ID is complex and involves many professional caregivers. For example, a person with an ID is in close contact with a personal care aide who helps him or her with their daily activities and medication, but also sees a dentist, a general practitioner, and a physical therapist. All caregivers use some type of health information system (HIS) to register their activities and to ensure reimbursement of care provided. This is not without consequence: according to Hanekamp et al. [11], care professionals in long term care in the Netherlands spend 35% of their time on administrative tasks, mainly in their respective HISs.

Typically, HISs are not only used for reimbursement but also to assist professionals and organizations to contribute to high-quality, efficient patient care [1, 34, 35]. Well-functioning HISs, are therefore an important building block in the care for people with an ID [36, 37]. In the context of our research HISs contain and manage (medical) information about the person with the ID. Other frequently used terms for HISs in ID care are: electronic health records, and electronic client (or patient) dossiers/records.

Because of the role of HISs in, amongst other, reimbursement [38], patient safety [39], communication amongst professionals [40] and the time spent on administrative tasks in HISs [41], it is essential to know about the HISs used in the complex care for people with ID. It is, however, unknown which HISs are used in ID care, how these systems cater to the needs of different groups of professionals, and what features are used and needed. Reports from multiple countries suggested that different HISs are not interoperable [42–44]. Despite several studies on HISs in other forms of care, such as hospital care and GP care [45–49], little research has been done on the problems professional involved in ID care experience with HISs, and on their requirements for HISs. Software developers and researchers need to know about tasks, problems and requirements of care professionals to develop better systems. Therefore, the objective of this study is to identify for ID care, the needs of care professionals by taking inventory of their experiences and wishes for HISs. To fulfill the objective, we set up the following research question: *In care for people with ID, what are the key care professionals' experiences with, and desired requirements for, HISs?* With the following sub-questions:

- Which HISs do care professionals for people with an ID use and for which purposes?
- How do care professionals for people with an ID assess their current HIS?
- Which features are care professionals for people with an ID missing in their current HISs?

To answer these questions, we performed a cross-sectional study in which Dutch care professionals working in ID care received a questionnaire informed by semi-structured interviews.

2.2 Methods

In the following, we describe the design of the online questionnaire, the recruitment of study participants, and data analysis.

2.2.1 Questionnaire construction

As preparation of questionnaire design and to assure thorough familiarity with this diverse care field, we interviewed from our personal and professional networks fourteen health care professionals who differed in terms of their profession but who all worked in ID care. The semi-structured interviews lasted about an hour on average and during site visits, interviewees also showed us their local HISs. All interviewees agreed with audio-recording. Transcriptions of the recordings were analyzed to identify routing requirements and in vivo coded (using Atlas.ti 8) to identify possible answer categories for the standardized questionnaire.

From the interviews and site-visits we learned that two types of HISs were distinguished in practice which differed with respect to users and aim: Electronic Client Dossiers (ECDs) and Electronic Patient Dossiers (EPDs). Care professionals in ID care mainly used the ECDs to register daily activities for care continuity. The EPDs were used for clinical purposes by, for example, dentists and ID physicians. We, therefore, built three routings in the questionnaire: one for the ECDs, one for EPDs, and one for care professionals who used both kinds of systems.

The in vivo coding of the transcripts yielded a total of 18 features and 9 problems with HISs in ID care. Because of interview limitations, we also included an 'Other, ...' answer option. The first concept of the questionnaire was pre-tested by means of a cognitive interview [50]. Based on this interview the questionnaire was improved, until after the third cycle no new issues appeared.

2.2.2 Recruitment of study participants

We recruited a wide range of health care professionals working in ID care as study participants via various Dutch professional associations and care providers. We were unable to obtain e-mail lists of the members of the professional organizations and the employees of the institutions due to Dutch privacy regulations. Therefore, we used tailored recruitment methods. Personal care aides, nurses, resident assistants (for assisted living), and sheltered workshop guides/job coaches received an e-mail from long-term care providers and their professional association 'Professional Association of Social Work Professionals'(BPSW). ID physicians were recruited by an e-mail from the 'Dutch Association of ID-physicians'(NVAVG). Remedial educationalists and behavioral therapists were recruited via an e-mail and the newsletter from their professional association 'Dutch Association for Remedial Educationalists and Behavioral Therapists'(NVO). Psychologists were recruited via a LinkedIn post from the 'Dutch Association for Healthcare Psychologists'(NVGzP). The 'Dutch Association for Psychiatry'(NVvP) published an invitation for participation on their intranet. Pharmacists were approached by e-mail through the 'Dutch Association of Pharmacists for People with an Intellectual Disability'(NVAPVG). Dentists were approached via the website, newsletter, and Facebook page of The 'Central Consultation for Special Dentistry', and the newsletter and Facebook page of 'Association for the Promotion of Dental Care for the Disabled'(VBTGG) . The 'Dutch Association of Physical Therapists for People with an ID'(NVFVG) invited their members via their newsletter.

2.2.3 Data collection procedure

After opening the online questionnaire, participants were requested to read information on the study's objective and give informed consent. The first question inquired after profession and HIS(s) used to determine routing. The questionnaire continued with questions on satisfaction, features, and problems related to the HIS(s) and concluded with questions on gender, the region of employment, and work experience. Data were collected using Qualtrics [51]. Data collection took place between September 2nd 2019, and October 15th, 2019. All questionnaires are available upon request from the corresponding author. All methods were carried out in accordance with relevant guidelines and regulations.

2.2.4 Data analysis

'Other, ...' answers were carefully read, and categorized by three researchers independently. Differences in classifications were discussed until consensus was reached. Then, new categories were added to the data file. We distinguished four groups of care professionals based on where they provide care, frequency (daily or not) and kind of professional training (academic or not): Daily care, Intellectual Disability physician, Mental Health and Development (MH&D), and Other care (see Table 2.1). Members of the Daily care group, although diverse in terms of profession, all work for a ID care provider. In the ID physician group, we also included ID physicians in training. The Mental Health and Development group works with people with ID on a daily basis, either within an ID healthcare provider or not. The remaining Other group is involved in care for, amongst others, people with an ID.

We investigated differences between these groups, and between EPDs and ECDs. With the non-parametric Kruskal-Wallis test we tested (with $\alpha = 0.05$) differences between the four groups on HIS' satisfaction. All statistics were obtained with R.

2.3 Results

2.3.1 Descriptives of respondents

Of the 328 respondents in Table 2.1, 275 also answered the last question (i.e. 16.2% drop-out). All respondents answered three or more questions. Nearly half of the respondents belonged to the Daily care group (n=144).

Table 2.1: The respondents per group of care professionals. Respondents were able to select one, or multiple professions.

Daily care	
Personal care aide	85
Nurse	45
Resident assistant at assisted living	32
Sheltered workshop guide, job coach	13
Other[a]	9
Total	144
ID physician	
ID physician	83
ID physician resident	8
Total	91
Mental health & Development	
Remedial educationalist	22
Psychologist	19
Behavioural therapist	5
Psychiatrist	4
Therapist, other	3
Other[b]	3
Total	54
Other care	
Pharmacist	15
Dentist	9
Physical therapist	5
Medical Doctor	5
Other[c]	5
Total	39
Total	328

[a]: Service coordinator, Functional manager, Policy officer, Extramural care professional, Trainer, care coordinator, Case manager, ECD expert, and Maternity aide.
[b]: Mental healthcare worker, Remedial teacher, and Prevention assistant
[c]: Dental laboratory technician, researcher, and three General practitioners

The respondents worked between one and 43 years in ID care, the median being 13 years (Table 2.2). Daily care professionals worked the longest in ID care with a median of 17.5 years. 80% of the respondents were women, only the ‘Other care’ group had a lower percentage, with 58%.

Table 2.2: Descriptive statistics on the respondents and systems.

		Group				
		Daily Care	ID Physician	MH&D	Other Care	Total[a]
Years worked in ID care	N	120	82	38	30	270
	Median	17.5	10	11.5	10	13
	Range	1.5-43	2-43	1-35	2-39	1-43
N systems mentioned	N	143	139	54	39	375[b]
N unique systems	ECD	12	8	13	4	20
	EPD	0	12	8	19	32
%Women	N	122	83	39	31	275
	Percentage	83	82	82	58	80

[a] The number of respondents from this table differs from Table 1 due to missings

[b] Some care professionals use two systems

2.3.2 Which HISs do care professionals for people with an ID use and for which purposes?

In total, 52 unique HISs were identified (Table 2.2), of which the majority (32) were EPDs. Table 2.3 shows the features used in ECDs (n=225) and EPDs (n=118), and satisfaction/dissatisfaction with that feature. Overall, Patient/Client administration was most frequently mentioned, followed by Reporting and Client treatment and Support registration. The majority of ECD users indicated other registrations, e.g. freedom of movement, and visitations. In the Appendix in additional Table 1 results were shown for each of the four different groups of care professionals.

Table 2.3: Respondents' use and satisfaction with the ECDs and EPDs features (N = Respondents, Likert scale; 1 = very dissatisfied, 2 = dissatisfied, 3= neither satisfied nor dissatisfied, 4 = satisfied, 5 = very satisfied)

	Users	**ECD**			**EPD**		
Feature	N	N	Mode	Range	N	Mode	Range
Patient/Client administration	238	149	4	1-5	89	4	2-5
Reporting	207	194	4	1-5	13[b]	5	1-5
Client treatment and support registration	194	183	4	1-5	11	4	2-5
Storage and document management	171	162	4	1-5	9[c]	4	2-4
Register medical patient information	120	10[a]	2	1-5	110	4	1-5
Financial administration and reimbursement	109	69	4	1-5	40	4	1-5
Calendar Management	101	36	4	1-4	65	4	1-5
Registration of consultation following struct.	96	-	-	-	96	4	1-5
Making of letters	92	-	-	-	92	4	1-5
Test results from specialist/lab	90	-	-	-	90	4	1-5
Registration of diagnoses	81	-	-	-	81	3	1-5
Communication between team members	72	67	4	1-5	5[a]	5	2-5
Prescribe medication	70	-	-	-	70	3	1-5
Client portal	68	67	3	2-5	1[a]	3	-
Electronic exchange of patient/client dossier	62	-	-	-	62	2	1-5
Medication overview	15	-	-	-	15	4	3-5
Medication surveillance	14	-	-	-	14[c]	4	3-5
Other registrations	172	169	4	1-5	3[a]	NU[d]	3-5

[a] Obtained from 'Other, ...'
[b] Answer option provided to Dentists only
[c] Answer option provided to Pharmacists only
[d] NU = Not Unique

2.3.3 How do care professionals for people with an ID assess their current HIS?

The groups differed on their satisfaction level with respect to their ECD in general (Kruskal-Wallis, p< 0.0001), with the Daily care group expressing higher levels of satisfaction than the other groups, in particular the Mental Health & Development group (Table 2.4). No evidence was found for different satisfaction levels with the general use of EPDs (Kruskal-Wallis, p=0.061). For general satisfaction with their EPD, all groups' modes were 'Satisfied'.

For the satisfaction with the suitability of ECDs for ID care a significant difference between groups was found (Kruskal-Wallis, p< 0.0001). Daily care professional's reports ranged from dissatisfied to very satisfied, with the most frequently occurring judgment 'satisfied' (Table Table 2.4). Based on range and mode, the ID physicians and Mental Health & Development group appear less satisfied with the suitability of the ECD for ID care. For EPDs, no evidence for differences between groups was found (Kruskal-Wallis, p= 0.12). The satisfaction with the suitable of EPDs for ID care appears rather low across all groups with the modes 'dissatisfied', and 'neither satisfied nor dissatisfied'.

Table 2.4: The care professionals' satisfaction with their system per group. Their frequency (N), Mode, and range of satisfaction (Likert scale, 1 = very dissatisfied, 2 = dissatisfied, 3= neither satisfied nor dissatisfied, 4 = satisfied, 5 = very satisfied)

			Group				
	System		Daily Care	ID Physician	MH&D	Other Care	Total
How satisfied or dissatisfied are you with your HIS in general?	ECD	N	143	63	40	4	250
		Mode	4	3	2	NU[a]	4
		Range	1-5	1-4	1-4	2-5	1-5
	EPD	N	-	77	14	36	127
		Mode	-	4	4	4	4
		Range	-	1-5	1-4	1-4	1-5
How satisfied or dissatisfied are you with the suitability of the HIS for ID care?	ECD	N	142	63	40	4	249
		Mode	4	3	3	3	4
		Range	2-5	1-5	1-5	3-5	1-5
	EPD	N	-	77	14	36	127
		Mode	-	2	2	3	3
		Range	-	2-5	1-4	2-5	1-5

[a] NU = Not Unique

When we go back to Table 2.3 for more detailed information on dis/satisfaction with features, we see that although the modes suggested satisfaction with the most frequently used features of ECDs, there was no consensus: satisfaction ranged from very dissatisfied to very satisfied. The satisfaction with the less frequently mentioned ECD features, suggested that some people have experienced serious problems with the registration of medical patient information.

A similar situation can be seen for EPD features: although modes suggest that users were rather satisfied with most features, there was no consensus at all. The modes 'neither satisfied nor dissatisfied' for the features 'registration of diagnoses' and 'prescribe medication' should be seen as an important signal: not only because these features are crucial to medical treatment and monitoring but also because they support medication safety. The lowest mode, dissatisfaction, was found for the electronic exchange of patient/client dossier.

The overall most frequently mentioned problems related to HIS design were 'hard to retrieve information from the system' (185 times), followed by 'difficult to exchange electronic client/patient records with other caregivers' (Table 2.5). Most other problems mentioned by over 100 respondents related to system reliability: slow, and unavailable systems, and updates that changed the system. Based on the 'Other. . . ' answers we added two new problem categories: Bad user interface, and User roles and permissions. Only 15 of the 277 care professionals indicated they had experienced no problems with one of their systems. (Table 2.5). See additional Table 2 in the Appendix for data on group level.

Table 2.5: Most frequently identified HIS problems, for ECDs (N=221 respondents) and EPDs (N= 111 respondents)

Problem	ECD	EPD	Total
Hard to retrieve information in system	124	61	185
Difficulty exchanging electronic client/patient records with other caregiver	76	73	149
System is slow	103	44	147
Having to work in multiple systems at the same time	65	70	135
System is unavailable	86	40	126
Updates change the system	80	26	106
Hard to exchange information with other systems within care institution	74	8	82
Primary care classification method not differentiated enough for ID care	-[b]	75	75
Bad user interface	5	1	6
User roles and permissions	3	0	3
Other problems[a]	4	1	5
No problems at all	10	5	15

[a] Problems that could not be classified into one of the above problems
[b] Not provided as an answer option in the survey for this system

2.3.4 Which features are care professionals for people with an ID missing in their current HISs?

Eighty-eight respondents suggested desired, missing, features. We categorized these into 24 missing features (Table 2.6). The most frequently mentioned missing feature was the link with other systems (20 times), followed by features to provide an overview of data in the system and to prescribe and monitor medication (12 times). Features that were reported missing by one respondent only were: Assign tasks, calendar management, financial management, taking history and making care-plan, open multiple dossiers at the same time, privacy robust procedures, suitability of system across disciplines, synchronization Windows and ECD, and 'other registrations'. Additional Table 3 in the Appendix shows the missed features per group of care professionals.

Table 2.6: The features the care professionals reported missing. ECDs N=51, EPDs N=37

Missing Feature	**ECD**	**EPD**	**Total**
Link with other systems	8	12	20
Providing overview	6	6	12
Prescription management and monitoring	5	7	12
Information exchange	4	6	10
No feature but problem	8	1	9
Uploading of files	3	5	8
External correspondence	4	3	7
Clinical notes management	1	5	6
Electronic prescription	2	4	6
Access to parts of system	2	2	4
Epilepsy module	3	1	4
Detailed reporting	1	2	3
Lab information	1	2	3
Data search and filter	0	2	2
Insult registration	0	2	2
Other	6	4	10

2.4 Discussion

2.4.1 Reflection on the results

To the best of our knowledge, this is the first study on HISs in ID care that included both medical and non-medical care professionals who all report on well-being and medical issues concerning the same individuals. We identified ECDs used by all groups involved in residential care and EPDs for medical care. Over 300 respondents reported a total of 52 unique HISs. Care professionals used different HISs for a variety of tasks ranging from administration to treatment registration and reporting. They also reported support registration and a client portal, features specific to ID care and illustrating complexity of this care setting. Information is scattered over different systems which challenges professionals in ID care.

Often people with ID express medical problems, like pain, in behavioral change [52]. This implies that medical staff may experience difficulties because they can't access the ECD and professionals in daily care may be able to do a better job if they have access to EPDs. Despite a reasonable overall satisfaction with HISs, nearly all respondents reported problems. Many problems related to HIS design (e.g. problems with the user interface), and reliability (slow, and outdated systems). Other problems stem from the organization of ID care, with multiple HISs, information exchange needs, and missing features.

2.4.2 Related work

Previous studies reviewed HIS satisfaction and adoption level for one user group, i.e. medical doctors or nurses in hospitals [45–47, 53–55] or in other care domains [48, 49, 56–60]. In general, these studies showed dissatisfaction with HISs, despite high adoption rates. Comparability between these studies depends on the national organization of healthcare, and is therefore limited(e.g. the GP as gatekeeper for specialized care, the presence of an ID-physician) [61, 62]. Comparability with our study is also limited because the care for people with ID is characterized by a combination of support and care services [63–65].

Other studies [40, 66–70] assessed the adoption and satisfaction with HISs based on features but not on what was missing. Generally speaking these studies found, just like ours, similar levels of satisfaction with the features.

A study in the ID care setting, considered governmental databases as information systems is Karimi et al. [71]. The features they distinguished are, therefore, types of data saved in these databases, and resemble outcomes of our features 'patient/client administration', 'reporting', and 'financial administration and reimbursement'. In their discussion, Karimi et al. [71] stressed the necessity of data exchange, which was also mentioned by our respondents.

Like ID care, geriatric care is a form of long-term care, but with better studies on HISs [72–74]. Wang and Biedermann [73] and Cherry et al. [74] primarily studied adoption and use of HISs, but not so much satisfaction. Sockolow et al. [72], conducted a survey on one HIS used in a geriatric community setting and concluded that although Philadelphian clinicians did not use the system as intended by the developers, they were satisfied with their HISs. We did not assess whether the care professionals used the systems as intended, since we did not study one HIS, but as it turned out, 52 HISs.

2.4.3 Strengths & weaknesses

The collaboration with nine professionals' associations and various institutions' willingness to participate shows the importance of this study. With their help, we could review a wide range of HISs used in ID care. Despite the limited role of GPs in care for institutionalized people with an ID we have attempted to gain the interest of multiple professional organizations for GPs, unfortunately they did not want to participate. Nonetheless we were able to review the most common GP HISs, mainly thanks to the Dutch ID-physicians. We believe our sample reflects the main professions responsible for the daily care of people with intellectual disabilities. Nonetheless, there are professions from which no representatives were included in our sample or which were underrepresented such as speech therapists, nutritionists and medical doctors other than ID physicians. However, problems with information exchange between these professionals and the respondents of our survey could be reported and, indeed, have been mentioned.

From the interviews, we derived our online questionnaire, including the set of features, which we tested and enhanced with cognitive interviews. We chose not to use a standardized set of features and criteria [75, 76], since these appeared too limited, and too coarse for ID care. The questionnaire contained some questions with a fill-in 'Other, . . . 'answer option. By discussing and synthesizing these answers, we experienced the added value of authors from different disciplines.

For the satisfaction questions, we adopted Likert-type scales despite their metric limitations, because they are familiar and therefor easy for respondents. Despite our carefully designed questionnaire, the drop-out rate was 16%. Fortunately our data was rich, because many and various respondents gave 'Other, ...' answers.

2.4.4 Practical implications

This study showed that the current EPDs are not satisfactory for the special needs in ID care. This indicates that in-depth knowledge of the specific ID care setting is necessary for the development of HISs, since it is more complex than other types of care. In the Netherlands, complexity might further increase due to the recently introduced legal obligation to give people with ID access to their own dossier [77]. This access is usually provided through a client portal which is also listed as a feature in Table 2.3. In practice, relatives often also have access to this system, which allows them to monitor care. Also, the legal representatives make the decisions about what happens to the data of the people with intellectual disabilities, if they cannot decide about these topics this themselves. This creates a complexity for system access and usage of the data. Furthermore, the large number of identified HISs (52) causes multiple interoperability problems.

This complexity, and the aforementioned problems with usability calls for a modular system, that serves as an ECD, EPD, and client portal, where ID modules with particular features can be removed and added as required. The modular system may reduce costs, because of it's decomposability.

A modular HIS for ID care should be based on standards for information exchange and a reference software architecture [78]. Furthermore, a reference architecture may help make data in ID care more suitable for research by combining privacy standards with the FAIR data principles [64].

2.5 Conclusions

In conclusion, answering the general research question, in care for people with ID a wide range of different HISs is used. Despite some satisfaction with the HISs, many problems remain to be solved. Future work will be the design of a reference architecture for modular HISs in ID Care.

Appendix

Additional table 1: (Table 3 by care professional groups): Respondents' satisfaction with the features they reported using (N = Respondents who use the feature, followed by Mode and Range; 1 = very dissatisfied, 2 = dissatisfied, 3= neither satisfied nor dissatisfied, 4 = satisfied, 5 = very satisfied)

	Users	**Daily care**			**Intellectual Disability Physician**			**Mental Health and Development**			**Other care**		
Feature	N	N	Mode	Range	N	Mode	Range	N	Mode	Range	N	Mode	Range
Patient/Client administration	238	94	4	1-5	84	4	1-5	36	4	1-5	24	4	2-5
Reporting	207	121	4	1-5	37	4	1-5	33	3	1-5	16[c]	5	1-5
Client treatment and support registration	194	108	4	1-5	38	3	1-5	35	2	1-5	13	4	2-5
Storage and document management	171	98	4	1-5	27	4	1-5	34	4	1-5	12[d]	4	2-5
Register medical patient information	120	2[a]	NU[b]	2-5	80	4	1-5	9	3	1-4	29	4	1-5
Financial administration and reimbursement	109	27	4	2-5	33	3	1-5	25	4	1-4	24	4	2-5
Calendar Management	101	19	4	1-4	44	4	1-5	21	3	1-4	17	4	2-5
Registration of consultation following structure	96	-	-	-	73	4	2-3	10	4	2-4	13	4	1-5
Making of letters	92	-	-	-	69	4	1-5	9	3	1-4	14	4	3-5
Test results from specialist/lab	90	-	-	-	73	4	1-5	7	3	1-4	10	3	3-5
Registration of diagnoses	81	-	-	-	68	3	1-5	11	4	1-4	2	NU[b]	4-5
Communication between team members	72	35	4	1-5	22	4	1-5	12	4	1-4	3[a]	NU[b]	2-5
Prescribe medication	70	-	-	-	48	3	1-5	-	-	-	22	4	1-5
Client portal	68	42	4	2-4	15	3	2-5	11	3	2-5	-	-	-
Electronic exchange of patient/client dossier	62	-	-	-	45	3	2-5	5	2	1-5	12	5	2-5
Medication overview	15	-	-	-	1[a]	3	-	-	-	-	14[d]	4	4-5
Medication surveillance	14	-	-	-	-	-	-	-	-	-	14	4	3-5
Other registrations	172	98	4	1-5	39	4	1-5	30	3	1-5	5[a]	NU[b]	2-5

[a] Obtained from "Other, ..."

[b] NU: Not Unique

[c] Answer option provided to Dentists only

[d] Answer option provided to Pharmacists only

Additional table 2: (Table 5 by care professional groups) Most frequently identified HIS problems, for each group of care professionals. MH&D = mental health and development

Problem	Daily Care	ID Physician	MH&D	Other Care	Total
Hard to retrieve information in system	60	80	30	15	185
Difficult to exchange electronic client/ patient dossiers with other caregiver	29	77	21	22	149
System is slow	59	58	18	12	147
Having to work in multiple systems at the same time	17	86	15	17	135
System is unavailable	55	47	18	6	126
Updates change the system	53	29	17	7	106
Hard to exchange information with other systems within care institution	36	24	12	10	82
Primary care classification method not differentiated enough for ID care	-[b]	64	6	5	75
Bad user interface	3	1	2	0	6
User roles and permissions	1	2	0	0	3
Other problems[a]	2	2	1	0	5
No problems at all	9	3	0	3	15

[a] Problems that could not be classified into one of the above problems

[b] Not provided as an answer option in the survey for this system

Additional table 3:(Table 6 by care professional groups): The features the care professionals reported missing. MH&D = mental health and development

Missing Feature	Daily Care	ID Physician	MH&D	Other Care	Total
Link with other systems	1	16	0	3	20
Providing overview	3	7	1	1	12
Prescription management and monitor.	1	8	0	3	12
Information exchange	1	7	1	1	10
No feature but problem	1	5	3	0	9
Uploading of files	3	3	1	1	8
External correspondence	0	5	2	0	7
Clinical notes management	0	5	0	1	6
Electronic prescription	0	5	0	1	6
Access to parts of system	0	4	0	0	4
Epilepsy module	1	2	0	1	4
Detailed reporting	1	2	0	0	3
Lab information	0	2	0	1	3
Data search and filter	0	1	0	1	2
Insult registration	0	2	0	0	2
Other	2	2	5	1	10

Study 2: Systematic literature review on health information systems

3

Background: Currently many healthcare systems are supported by an increasing set of Health Information Systems (HISs), which assist the activities for multiple stakeholders. The literature on HISs is, however, fragmented and a solid overview of the current state of HISs is missing. This impedes the understanding and characterization of the required HISs for the healthcare domain.
Methods: In this article, we present the results of a Systematic Literature Review (SLR) that identifies the HISs, their domains, stakeholders, features, and obstacles.
Results: In the SLR, we identified 1340 papers from which we selected 136 studies, on which we performed a full-text analysis. After the synthesis of the data, we were able to report on 33 different domains, 41 stakeholders, 73 features, and 69 obstacles. We discussed how these domains, features, and obstacles interact with each other and presented suggestions to overcome the identified obstacles. We recognized five groups of obstacles: technical problems, operational functionality, maintenance & support, usage problems, and quality problems. Obstacles from all groups require to be solved to pave the way for further research and application of HISs.
Conclusion: This study shows that there is a plentitude of HISs with unique features and that there is no consensus on the requirements and types of HISs in the literature.

This chapter is based on an article published in an international journal:

Tummers, J., Tekinerdogan, B., Tobi, H., Catal, C., & Schalk, B. (2021). Obstacles and features of health information systems: A systematic literature review. Computers in Biology and Medicine, 137, 104785.

3.1 Introduction

Nowadays the healthcare sector is becoming increasingly dependent on the supporting information systems. Currently, for example, almost every registration happens digitally and digitization in healthcare is rising [79]. Only a few decades back, the first transition was made from the paper-based record to the electronic health record [80], today we discuss using techniques such as blockchain to exchange health data [81]. The first digitization arrived at academic medical facilities [80, 82], but since then also other domains of the healthcare sector use their information systems. For example, a General Practitioner (GP) uses a GP information system, a pharmacist a pharmacy information system, and a laboratory technician a laboratory information system. All these systems can be categorized as Health Information Systems (HISs).

Typically HISs assist healthcare organizations in processing data, information, and knowledge in order to contribute to high-quality, efficient patient care [1, 34, 35]. The HIS is considered as a building block of the complete health system, and therefore a well-functioning HIS is a vital aspect for delivering excellent care and receiving reimbursement for the care given [36, 37]. This importance also emerges from the following statement of the WHO [36]: *"A well-functioning health information system is one that ensures the production, analysis, dissemination and use of reliable and timely information on health determinants, health system performance and health status."* For this article, we consider a comprehensive HIS from all different domains of the healthcare sector.

Whereas in the past registration was often done on a dedicated desktop computer, today it is also increasingly done on other platforms such as mobile phones, tablets, or a central computer server [83–85]. Due to the differences between healthcare domains and deployment methods of HISs, each HIS has unique characteristics, which we define as system features. HISs used to have the sole function of keeping track of medical history, but the range of features is currently much wider. All features together define the functionality of the HIS. Widely used features of HISs are, for example, clinical notes, medication registration, and electronic health record management [86–88].

In practice, several obstacles can be identified related to the development, usage, and maintenance of these HISs. Similar to the features of HISs, these obstacles have been reported in different studies. Although HISs have been considered in multiple studies, it remains unclear which features are provided by an HIS, which are the obstacles of HISs, and which stakeholders and healthcare domains they support. The objective of this study is to use a systematic literature review (SLR) to assess the current state of HISs. By current state, we mean mapping who is using the systems (stakeholders), where they are using it (domain), how they are using it (deployment), what they are using it for (features) and what they are facing in this use (obstacles).

Most previous literature reviews focus on one or two obstacles, such as privacy [89, 90] , interoperability [81, 91], or quality [92, 93]. These reviews attempted to find solutions to these obstacles based on the articles identified. However, a literature review that identifies and quantifies all the different types of obstacles is not currently available. Such a literature review could help provide an overview of the wide range of obstacles faced by the various stakeholders. The study can be of benefit for both practitioners and researchers. Practitioners could identify the obstacles and take these into account for developing the required system. Researchers can focus on the identified obstacles and challenges and thus further advance the research in this domain.

We use the guidelines of Kitchenham et al. [94] for performing SLRs, which were developed for the software engineering domain. Therefore, this is not a generic narrative literature review but focuses on a limited number of aspects of HISs that we want to quantify in this study. We begin by presenting the review protocol in Section 2, and the paper continues with the results in Section 3.3. Section 3.4 presents the discussion, and finally, Section 3.5 provides the conclusions.

3.2 Review protocol

For the SLR, we followed the guidelines presented by Kitchenham et al. [94] in which guidelines from, amongst others, the Cochrane Collaboration [95], are adapted for software engineering.

Our review protocol is presented in Figure 3.1. It started with the definition of the search strategy to construct the query; this was an iterative process until we got a search string that we found sufficient. Then, in the second step, a set of study inclusion and exclusion criteria was identified. To exclude low-quality studies, we designed the quality assessment method in step three. In the fourth step, we designed a data extraction form with which we could extract the elements (healthcare domains, stakeholders, features, and obstacles) needed to answer the research questions. In the fifth step, a synthesis method was determined to summarize the different text elements identified into results in the fifth step.

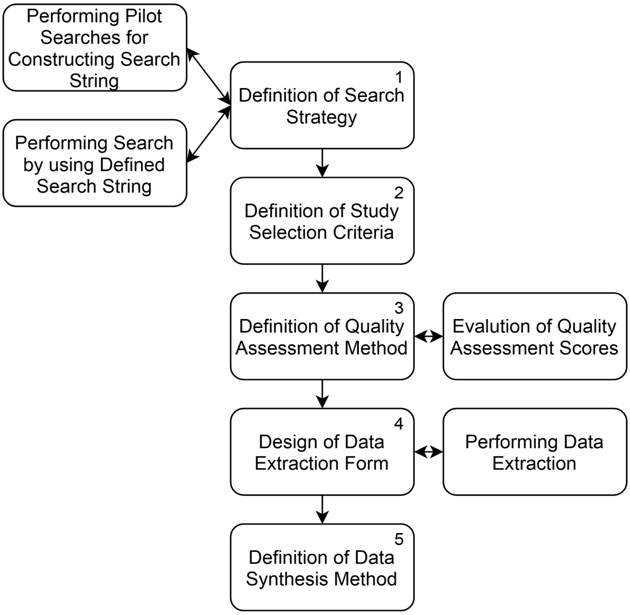


Figure 3.1: Review protocol for our SLR. Adapted from Kitchenham et al. [94]

3.2.1 Search strategy

We performed an automated search for full papers in three bibliographic databases: Web of Science (the most general library), IEEE Xplore (that aims at IT and computer sciences), and PubMed (that focuses on health and biomedical sciences) using standard filtering options to search for articles in computer science, medical information system, and medical informatics. The search query consisted of four building blocks (Table 3.1) and was performed on the title and abstracts.

Table 3.1: The search query used for this SLR. The syntax presented below was used for Web of Science, the other two databases required a slightly different syntax but contained the same building blocks

Building Block	Syntax	
Information technology	"Information Technolog*" or ICT or "Information and Communication Technolog*" or "Computer system" or "Information System" or Informatics	AND
Information System	"electronic health record" OR "Electronic medical record" OR "Electronic patient record" OR "patient record system" OR "Clinical decision support system" OR "Computer Patient Record" OR "Patient care information system" OR "Electronic client record" OR ((Health* OR Medical OR Patient OR Client) NEAR/4 (Manage* OR Informat*)) NEAR/4 (Software OR System* OR Tool OR Platform)	AND
Healthcare domain	Health* OR "General practitioner" OR "Family practice" OR Hospital or Physi* OR "Primary care" OR Paramedic OR Disease OR Prevent* OR care OR cure OR GP OR doctor OR dent* OR medic* OR pharma* OR psycho*	AND
Characteristics	Feature$ OR Function* OR Module$ OR Obstacle$ OR Problem$ OR Demand*	

The first three building blocks were used to ensure articles related to ICT and computer science on information systems used in health care. The last block, *"Characteristics"*, was added to make sure that the articles actually describe an HIS. The final search for papers published in 2009 – 2020 was performed on May 27th, 2021. The search yielded a total of 1480 articles, and after removing duplicates 1340 unique articles remained (see Figure 3.2).

3.2.2 Study selection criteria

We split the study selection criteria into inclusion and exclusion criteria (Table 3.2). These criteria were developed in an iterative process involving all authors and a random sample of 20 abstracts. We excluded studies that met any of the nine exclusion criteria and included studies that met inclusion criteria only. Two authors followed this strategy on a randomly chosen set of 23 abstracts, yielding an inter-rater agreement of 78%. If the authors disagreed, the abstract was discussed until consensus was reached. After applying the selection criteria on abstracts, 281 articles (see Figure 3.2) remained for which the full text was sought.

All articles could be retrieved, and in the process of reading the full papers, another 118 articles were discarded based on the selection criteria from Table 3.2, which left 162 articles for the quality assessment.

Table 3.2: Study exclusion and inclusion criteria

No.	**Exclusion Criteria**
EC1	The abstract does not describe one or multiple Health Information Systems
EC2	Study using an HIS, but study not targeted on an HIS
EC3	Health Information System for animals
EC4	Other than survey studies within an HIS as PROMIS
EC5	Papers without full text available
EC6	Papers not written in English
EC7	Duplicate publication from multiple sources
EC8	Papers do not relate to the health sector
EC9	Literature Review papers
No.	**Inclusion Criteria**
IC1	Study that develops one or multiple HIS
IC2	Study that reflects one or multiple HIS
IC3	Study that develops a new module for an HIS
IC4	Study that reflects on an module of an HIS
IC5	The system is actually implemented or is a prototype under development

3.2.3 Study quality assessment

The third step from the review protocol was the quality assessment of the 162 remaining articles. The quality assessment criteria (Table 3.3) were adapted from Kitchenham et al. [94]. For each of the eight criteria up to 1 point could be obtained (1 point for a fully met criterion, ½ for partly met, and 0 for not met at all), yielding quality scores between zero and eight points. For example, if the study's aim was clearly stated in the introduction (anticipated place), a full point (1.0) was awarded, and if the study's aim was not expressed, no point was awarded. If the aim was vaguely specified or not at the expected location, a half-point (0.5) was granted. These assessment criteria were the same or almost the same as those in [96–100].

Table 3.3: Quality Assessment Criteria

No.	**Question**
Q1	Are the aims of the study clearly stated?
Q2	Are the scope and context and research design of the study clearly defined?
Q3	Are the variables in the study likely to be valid and reliable?
Q4	Is the research process documented adequately?
Q5	Are all the study questions answered?
Q6	Are the negative findings presented?
Q7	Are the main findings stated clearly? regarding creditability, validity, and reliability?
Q8	Do the conclusions relate to the aim of the purpose of the study? Reliable?

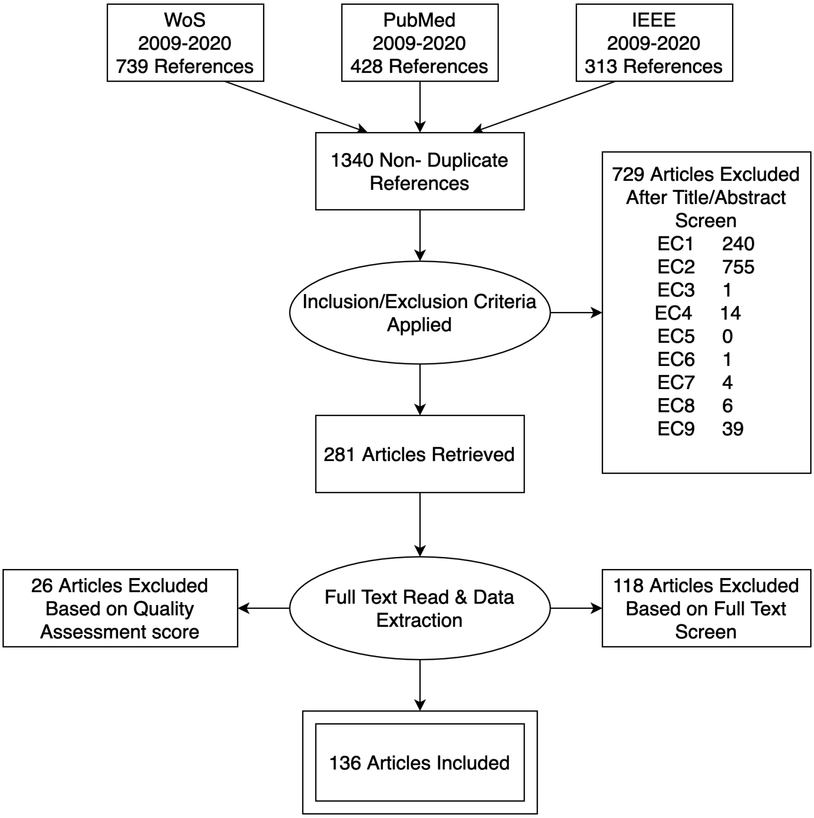
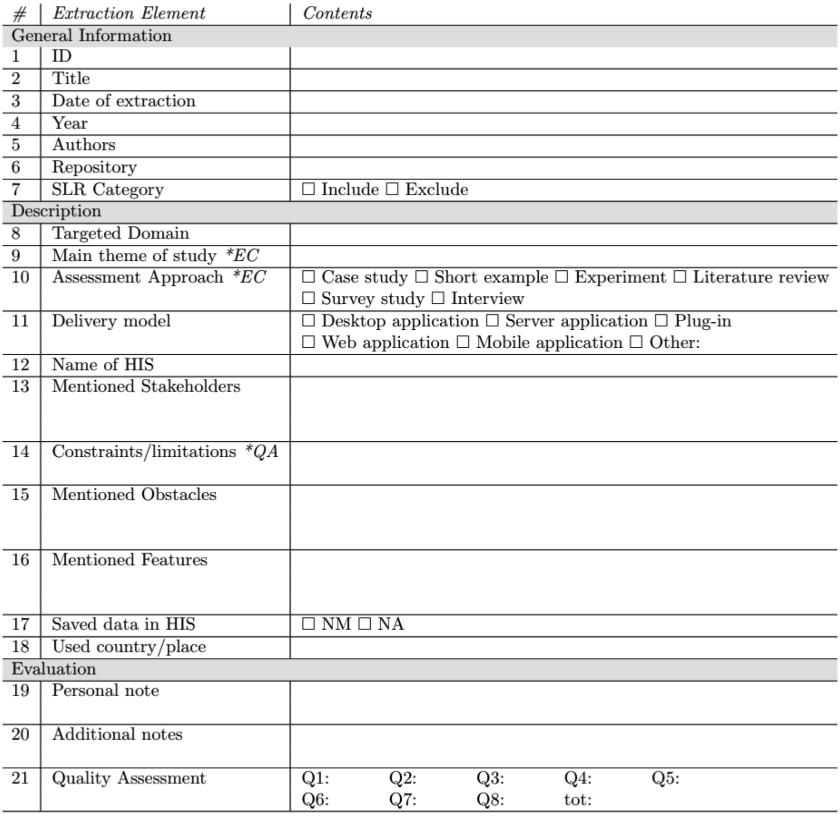


Figure 3.2: PRISMA statement flow diagram [101]

To be included, the articles required a score of four or higher, resulting in an analytic set of 136 articles (see Figure 3.2). The distribution of the quality scores is depicted in Appendix 3.5.

3.2.4 Data extraction

The development of the data extraction form was an iterative process where initially, the form converged after about a dozen papers into the form depicted in Figure 3.3 and was used for all 281 full-text articles. If a primary study mentioned one or more elements of the data extraction form, we wrote them down. We did not look for an exact match between article and form. For example, what we call targeted domain was amongst others called context, scope or sector in articles, and if an article listed actors to refer to stakeholders, we entered the list of actors in the form, maintaining the original wording. All elements from the data extraction form are explained in Appendix 3.5.



#	*Extraction Element*	*Contents*
General Information		
1	ID	
2	Title	
3	Date of extraction	
4	Year	
5	Authors	
6	Repository	
7	SLR Category	☐ Include ☐ Exclude
Description		
8	Targeted Domain	
9	Main theme of study **EC*	
10	Assessment Approach **EC*	☐ Case study ☐ Short example ☐ Experiment ☐ Literature review ☐ Survey study ☐ Interview
11	Delivery model	☐ Desktop application ☐ Server application ☐ Plug-in ☐ Web application ☐ Mobile application ☐ Other:
12	Name of HIS	
13	Mentioned Stakeholders	
14	Constraints/limitations **QA*	
15	Mentioned Obstacles	
16	Mentioned Features	
17	Saved data in HIS	☐ NM ☐ NA
18	Used country/place	
Evaluation		
19	Personal note	
20	Additional notes	
21	Quality Assessment	Q1: Q2: Q3: Q4: Q5: Q6: Q7: Q8: tot:

Figure 3.3: The data extraction form used in this study. NA = Not applicable, NM = Not mentioned

With the form, we identified that authors mentioned a total of 114 different healthcare domains, 437 stakeholders, 1926 features, and 982 obstacles.

3.2.5 Data synthesis

The synthesis for this study was done for the stakeholders, healthcare domains, features, and obstacles identified in the analytic set of 136 studies. For each of the elements (Healthcare Domains, Stakeholders, Features, and Obstacles) the following protocol was followed. After identification of duplicates and spelling dissimilarities, (near) synonym words were clustered into categories using a bottom-up strategy. The first author categorized all elements. The other authors categorized at least one element. The final categorization presented in this paper was a consensus decision based on plenary discussions. After the categorization, we got a total of 33 healthcare domains, 41 stakeholders, 73 features, and 69 obstacles.

3.3 Results

This section begins with a description of the 136 papers included in the analytic set and continues with the results for each research question. The references of the 136 included studies are listed in the Appendix.

3.3.1 Analytic set description

We included studies from the last twelve years; each year yielded at least 5 articles, the maximum being 21 articles from 2020 (for more detail see Figure 3.4). The most frequently occurring journal was the International Journal of Medical Informatics, with 23 occurrences. On the second place was BMC Medical Informatics and Decision Making closely followed by the journal of the American Medical Informatics Association, with fifteen and thirteen occurrences respectively. Forty-six journals only appeared once in the analytic set. The frequency distribution over the journals can be found in Table 3.4.

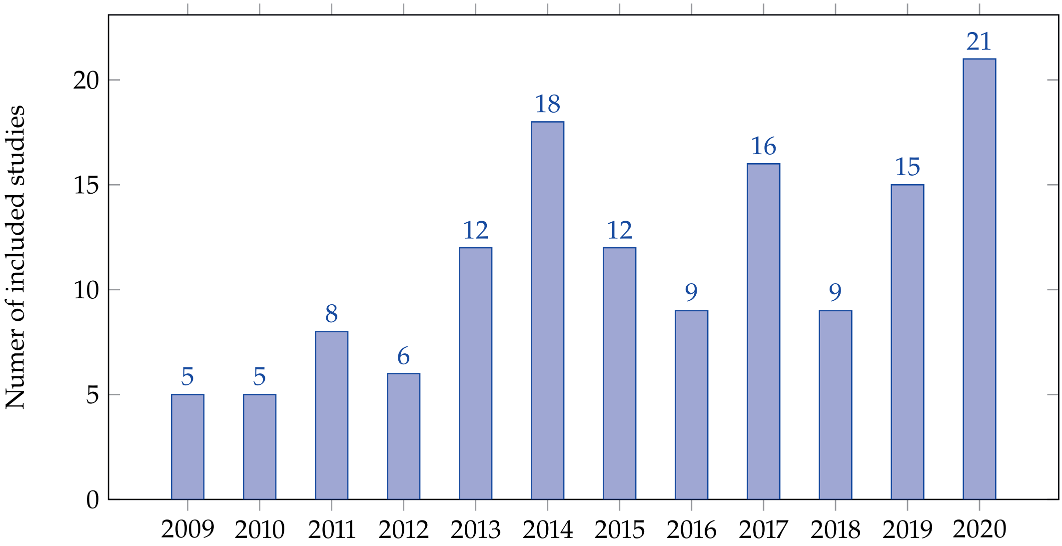


Figure 3.4: The year-wise distribution for the 136 studies in the analytic set.

Table 3.4: Journals that published two or more of the 136 studies in the analytic set.

International Journal of Medical Informatics	23
BMC Medical Informatics and Decision Making	15
Journal of the American Medical Informatics Association	13
Applied Clinical Informatics	7
Journal of Medical Systems	6
Methods of Information in Medicine	4
Health Informatics Journal	3
IEEE Access	3
JMIR Medical Informatics	3
CIN-Computers, Informatics, Nursing	2
Computer	2
EGEMS (Washington DC)	2
IEEE Journal of Biomedical and Health Informatics	2
Journal of Biomedical Informatics	2
Studies in Health Technology and Informatics	2

3.3.2 Which HISs are described in the literature?

In total, 128 unique HISs could be identified. The most frequently mentioned HIS was EPIC, but the vast majority appeared only once. Typically there seems to be a relation between the HISs and the healthcare domains, although some HISs seemed to have a broader scope and appeared in multiple healthcare domains. Different ways to cater to specific needs of functionality were observed: e.g. a mobile application to record blood sugar measurements, or a desktop application if much text needs to be entered. Most studies did not mention the deployment model of the HISs, others investigated multiple HISs and provided few details only. We identified that 28 studies mentioned a Web/Client-server application as the deployment model, eleven studies a Mobile application, and only seven studies mentioned the Stand-alone/desktop application as the deployment model.

3.3.3 Healthcare domains

We identified 33 different healthcare domains for the HISs (Table 3.5), of which some domains were location-specific (e.g. *hospital*) and others disease-specific domains such as *ophthalmology*. *Hospital* occurred most frequently followed by the *primary care* domain. In the articles from 2020, COVID-19 was mentioned three times as *infectious disease* domain.

Table 3.5: The identified healthcare domains from the analytic set of 136 studies and the number of articles in which they occur

Hospital	49	Allergy and immunology	2
Primary care	20	Brain disorder	2
Pediatrics	8	Cardiology	2
Infectious disease care	7	Community health	2
Laboratory	5	Nursing	2
Medication	5	Surgery	2
Outpatient care	5	Telehealth	2
Radiology	5	Alternative medicine	1
Diabetes care	4	Care for homeless	1
Care for chronically ill	3	Emergency care	1
Dentistry	3	Geriatric care	1
Maternal-fetal medicine	3	Ophthalmology	1
Mental health	3	Public health unit	1
Neurology	3	Rare diseases	1
Oncology	3	Telemedicine	1
Palliative care	3	Transmural care	1
Pulmonology	3		

3.3.4 Stakeholders

We identified 41 different stakeholders in the data, which we grouped according to their in/direct use or technical involvement in the HISs (Table 3.6). The largest group were the direct users of the HISs (mentioned in 114 articles). Indeed, *Physician* without further details and *Nurse* were the most frequently mentioned stakeholders. Seventy-eighth of the articles mentioned indirect system users who are influenced by the system, but do not utilize the system themselves such as *Patients*. Forty-five articles referred to technical stakeholders, such as *IT staff*. Articles related to the *hospital* domain had a focus on direct system users, whereas in-direct system users such as *patients* were mentioned in relatively few papers. Articles in the *primary care* domain identified a relatively large number of stakeholders and often represented opinions and experiences of direct users, in-direct users, and IT staff.

Table 3.6: The identified stakeholders from the analytic set of 136 studies and the number of articles in which they occur

Direct system users	Physician	84
	Nurse	63
	Medical specialist	43
	Administrative Staff	29
	Pharmacist	19
	Laboratory technician	16
	Therapist	14
	Medical assistant	13
	Regulator/policy maker	9
	Resident physician	7
	Secretary	7
	Educator	6
	Healthcare consultant	6
	Data clerk	5
	Dentist	5
	Receptionist	4
	Social worker	3
	Biomedical engineer	2
	Hospice staff	2
	Alternative medicine practitioner	1
In-direct system users	Patient	44
	Healthcare manager	30
	Researcher	13
	Patients family/relatives/representatives	5
	Student	5
	Counselor	2
	System administrator	2
	Healthcare association	1
	Insurance company	1
	National expert	1
	Health promotion worker	1
Technical stakeholder	IT-staff	22
	Healthcare informatician	18
	System developer	12
	Software vendor	7
	Statistician	4
	Technical staff	3
	IT expert	3
	System manager	1
	Technical writer	1
	External system	1

3.3.5 Features

We could identify 72 unique categories of features based on 128 of the 136 papers (see Table 3.7). The data on features differed in the level of abstraction due to the bottom-up approach we applied: they range from general features such as *Quality control* to specific functionalities such as *Death registration*. We then grouped the features into two types: general features, and sector-specific features. General features such as *Documentation* and *Reporting* relate to information systems in general, not necessarily to HISs, and were listed in 123 papers out of 136 papers. In addition, 125 papers mentioned sector-specific features. We observed that articles aiming at a specific medical domain, such as *radiology* or *neurology*, included sector-specific features, for example, related to medical imaging, without ignoring the general features.

Table 3.7: Identified features from the analytic set of 136 studies and the number of articles in which they occur

Sector specific		General features	
Medication recording	68	Reporting	57
Patient record	62	Order management	47
Lab test results	54	Reminders and alerts	45
Clinical decision support and guidelines	47	Administration and finance	43
Diagnosis/ clinical assessment	47	Video and image analysis	40
Patient tracking and monitoring	41	Documentation	39
Clinical notes	37	Appointments and scheduling	33
Treatment planning	33	Recording demographics	28
Recording vital signs	32	Problem list	27
Laboratory functionality	31	Internal communication	27
Specialist care feature	28	Data visualization	20
Medication prescription	26	Data import and export	20
Disease monitoring	25	Data and record exchange	20
Patient admission and registration	24	Staff and patient education	17
Radiology management	24	Data storage	17
Patient health status registration	22	External communication	15
Allergy recording	19	Data and system integration	14
Patient care overview and summaries	19	Security and risk management	12
Immunization and vaccination registration	16	Inventory management	11
Making discharge summaries	15	Data search	10
Consultation documentation	14	Workflow support	9
Referrals	14	Quality control	9
Patient portal	14	Authentication	7
Pharmacy functionality	12	Task management	6
Recording blood values	10	Remote access	6
Medical forms and questionnaires	9	Human resource mgmt.	6
Medical data analysis	8	Evaluation and benchmarking	5
Recording symptoms	8	Sensor management	3
Care coordination	5	Voice control	2
Clinical measurements	5	Setting goals	2
Food management	5	Prognosis	2
Death registrations	4	Help function	2
Telehealth	4	To do list	1
Visit management	3	Data sharing	1
Family planning	1		
Informing patient and family	1		
Lifestyle suggestions	1		
Pain recording	1		

3.3.6 Obstacles

We obtained a set of 69 obstacles, presented in Table 3.8. Most identified obstacles relate to missing features of the HISs and limited use of the system. We grouped the obstacles into five different groups: Technical problems, Usage problems, Quality problems, Operational functionality, and Maintenance & support. Technical problems include obstacles related to the installation of HISs, hardware availability, and network speed and availability. These technical problems are often a problem of the mediocre design of HISs. Often the root of these problems lies in the inadequate infrastructure of information and communication technologies.

Many usage problems such as poor system usability and problems with user training were identified. Fruther, several articles described a low adoption rate of HISs, leading to limited use. This is probably due to issues such as low user satisfaction and lack of limited education of stakeholders, which leads to the system being used less. Quality problems are often related to poor data quality but also to the poor quality of the HIS itself. Some studies even reported the loss of data and the failure of systems, which can affect patient safety. Many of these quality problems came from poor interoperability between different systems and the data in these systems. Obstacles related to the operational functionality of the system often resulted from missing features or problems with specific features of the system. This often led to users having to use multiple systems. Furthermore, the fit with the clinical workflow was not always optimal; often there was a different sequence here, which again could lead to taking away from the patient's attention.

Maintenance & support obstacles were reported to come from a lack of professional support and poor system maintenance. Often, care professionals found communication with the IT department to be poor to nonexistent, and furthermore, the documentation of the HISs was often not up to par either.

Nearly two out of three papers mentioned obstacles related to Technical problems (81 out of 136 papers) or Usage problems (80 papers). Obstacles related to Quality problems, Operational functionality, and Maintenance & support were mentioned in 67, 66, and 47 papers, respectively. Most obstacles seem to be generic and not related to a certain domain. Furthermore, we could not discover any connections between the obstacles, stakeholders, and features.

Table 3.8: Identified Obstacles from the analytic set of 136 studies and the number of articles in which they occur

Technical problems	Poor interface design	27
	Lack of standards	24
	Poor security	19
	Lack of data and system integration	18
	Hardware/Power problems	16
	Poor privacy	14
	Poor data and system availability	13
	Performance problems	11
	System limitations	10
	Network problems	10
	Poor hardware availability	10
	Poor data exchange	7
	Poor system accessibility	7
	Lacking infrastructure	6
	Data and system reliability	6
	System down	6
	System installation problems	2
	Poor scalability	2
	Problems with data & system storage	1
Operational func.	Missing features	47
	Bad fit with clinical workflow	17
	Problems with specific features	14
	Captures attention away from patient	12
	Requirement of multiple systems	6
	Work in unique & specific setting	5
	Poor working environment	4
Maintenance & Support	Lacking professional support	25
	High system costs	23
	Poor communication	11
	Legal and bureaucratic problems	9
	Lack of help and documentation	7
	Poor system updates	5
	Low trust in supplier	3
Usage problems	Limited use	36
	Time consuming to use	21
	Lacking user training	21
	Poor system usability	21
	High system complexity	20
	Uneducated users	15
	Data and system inefficiency	13
	Duplicate documentation	10
	Manual work	7
	Incorrect usage	6
	Navigation issues	6
	Low trust in system	6
	Low user satisfaction	5
	User disagreement with system	4
	Finding data	4
	Meeting user needs	4
	Information overload	2
	Lack of awareness	2
	Maximum use of system to usage	2
	Low system usefulness	2
	Hygiene problems	1
	Use of free text for registrations	1
Quality problems	Lack of data and system interoperab.	24
	Low data quality	21
	Poor system development	12
	Data and system inconsistencies	11
	Data loss	10
	Faulty system	9
	Data input/output/Propagation	8
	Poor data integrity	4
	Low system accuracy	4
	Poor patient safety	2
	Fragmented data	2
	Medical error	2
	Poor system natural language	2
	Data interpretability	1

3.4 Discussion

To the best of our knowledge, this is the first SLR for HISs that identified the current state of HISs by focusing on the healthcare domains, stakeholders, features, and obstacles. In this discussion, we critically reflect on the results, compare this study with related work, and discuss possible threats to validity.

3.4.1 Critical reflection on the results

Following the guidelines of Kitchenham et al. [94], we identified 136 primary articles from which we extracted data to identify the current state of HISs. A large percentage of the articles focused on HISs for hospitals and related domains, such as surgery and pathology. The dominance of HISs for hospitals probably influenced the distribution of categories of stakeholders and features we report. We think this influence is much smaller in the obstacles because the only health sector-specific category appears to be Usage problems.

The stakeholders we identified were mainly medical; not many different technical stakeholders were identified. This may be due to the dominance of medical informatics journals in our analytic set (see Table 3.4). Researchers or developers of HISs can use this list of stakeholders as input for stakeholder analysis, using a mapping technique such as that of Mendelow [102]. In the relatively few articles that mentioned delivery method, the web application was most often mentioned. This is a contradiction with practice, where still a lot of stand-alone applications are used, although the use of smartphones in healthcare is rising [103]. This contrast between the state-of-the-practice and the state-of-the-art may have to do with health care providers' concerns about hygiene and data leakage [104].

A better design of HISs may help overcome many of the obstacles related to technical problems. For example, one could use the guidelines from Zahabi, Kaber, and Swangnetr [105] or the standards from ISO 9241 part 12[106] to develop user-friendly and secure user interface for HISs. Other technical problems can often be overcome by using the right architecture for the right deployment. For example, in developing countries with frequent power and/or internet outages, having a desktop computer in combination with a web application is not very useful. A mobile application on the other hand, with a (temporary) local storage, can also function during power and/or internet outages. [107] The obstacles in the Usage problem group can often be overcome by more and better education and training for users of HISs. Younge, Borycki, and Kushniruk [108] reviewed multiple training methods for end-users with HISs and concluded that it is most effective to use a combination of training approaches and provide and continue training during implementation, orientation, and post-implementation. This can lead to the users appreciating the systems more and increasing the efficiency of using the HIS.

The obstacles related to the Quality of HISs and their data can be a significant barrier for further adoption and use, including the application of techniques such as machine learning. In order to ensure interoperability with other systems, standards such as HL7 FHIR should be used [109]. Furthermore, the diagnoses in the systems should also be standardized using codes such as ICPC-2 or ICD11 [110, 111]. Developers of HISs can use one of the many standards for healthcare data, such as reviewed in Schulz, Stegwee, and Chronaki [112], although there is minimal international consensus on the use of these standards.

Multiple obstacles are related to (missing) operational functionality of HIS, which might be because of the rapidly changing and increasing complexity of care. An overview of the available features helps to overcome problems related to the operational functionality. Additionally, it is important to ensure that the HISs integrate well with the clinical workflow. This can be done by a workflow assessment in multiple steps, as was done for example in [113].

Problems related to maintenance and systems support can be solved by better assistance from healthcare organizations, as well as from software vendors. According to Dehaghani and Hajrahimi [114], proper maintenance is key to the lifespan of HISs and it is crucial to determine the costs of maintenance in advance so that the costs are not getting out of hand. Good cooperation between the healthcare organization and the software supplier is key to solve the problems related to maintenance and support while keeping the costs under control. Furthermore, each HIS must have solid documentation and a manual that is understandable to the healthcare professional. This documentation must meet several requirements, as described in the IEEE/ISO/IEC 26511-2018 standard.

We foresee that the development of HISs in terms of features may go in two directions. There could be complete systems that contain almost all features, or there could be a switch to a platform structure that works with plug-ins from different developers. In order to move towards a platform structure, it is important to have a clear view of the architecture of HISs. A reference architecture for HISs would support HIS design by serving as a guideline. To the best of our knowledge, a reference architecture is, however, not yet available for HISs.

In this paper we scrutinized articles on the operational usage of HISs and suggest the importance of the yet lacking reference architecture. Future work will, therefore, involve the creation of a reference architecture for HISs based on domains, stakeholders, features, and obstacles identified.

3.4.2 Related work

Other reviews on HISs [115, 116] had a different focus or goal compared to this paper. Rahimi and Vimarlund [116] reviewed methods used to evaluate HISs but did not focus on the HISs themselves. They did indicate several obstacles that we also identified to user satisfaction, such as timeliness and usability, and costs of HISs. Sligo et al. [115] reviewed the literature on HISs project planning, evaluation, and implementation in their study on HIS influence on organizational change. Commonalities between this review and ours are mainly on obstacles, for example limited use, and the lack of trust in the IT capability of the supplier. These reviews [115, 116] studied the interaction with the system, and consequently identified many stakeholders' concerns, whereas we systematically looked at the system's side.

There are many more reviews (than [115, 116]) for specific health care domains [117–119]. This means that the literature on HISs is fragmented which impedes the understanding and characterization of HISs for the healthcare sector. A solid comprehensive overview of the current state of HIS was missing. E.g. Ball [118] presented perspectives on problems and prospects for hospital information systems. The article by Janett and Yeracaris [119] presents challenges with electronic medical records in U.S. primary care. Like our article, they present the lack of interoperability and missing standardization as an obstacle for further system development. All kinds of stakeholders, features, and obstacles presented in Ball [118] and Janett and Yeracaris [119] were also found in our study, but our study had a broader scope and, therefore, presented more diverse stakeholders, features, and obstacles.

Some articles do not give a complete overview of HISs but focus on one aspect only. For example, [120] does a stakeholder analysis to identify users of HISs. They identify only three major types of stakeholders, which they further detail in use cases: Physicians, clinicians, and patients, which are also present in Table 3.6 with more detail. Kumar et al. [121] identified research gaps in routine health information systems by means of a literature review for low- and middle-income countries. From 316 studies they identified obstacles related to data quality, data use, and system design. They present several related obstacles but do, unfortunately, not present them in a structured manner, which makes comparison with our article difficult.

Kruse et al. [122] performed an SLR with 55 primary studies on how electronic health records support population health. They identified a set of facilitators and barriers for the adoption. These barriers show many similarities with our obstacles, such as missing data, interoperability, productivity loss, and complex technology. They suggest more comprehensive standards for the interoperability of electronic health records. A study that mentions the key features of a European Union health information system is Rosenkötter et al. [123] .Unfortunately, they have a different notion of feature as in our study. The focus is on the most important aspects for developing a new system; for example, a feature in this study is "Sustainabilty". The four main functions of HISs according to the WHO [124] are data generation, compilation, analysis and synthesis, and communication and use. These functions are performed by the features that are in Table 3.7 under the "general features" group. Interestingly the WHO does not mention functions that are in our "Sector-specific" features group.

Papers which describes the challenges of HISs are Ngafeeson [125] and Roehrs et al. [126]. Ngafeeson [125] mention IT adoption as a challenge, but otherwise group the problems into four groups: related to the technology itself, the healthcare setting, the users of the system, and the regulatory environment. These challenges are all reflected in the obstacles in Table 3.8, where we provide more detail on the obstacles. The SLR with 48 primary studies from Roehrs et al. [126] on personal health records identifies many challenges and open issues for personalised health records (PHR). Furthermore, they also identify 20 data types in personal health records, which show many similarities with our identified features, such as demographics, documents, prescriptions, and scheduling. Subsequently, they provide an analysis on the architecture types of personal health records and list eighteen standards.

3.4.3 Addressing Threats to Validity

The main threats to validity for any SLR are publication bias, selection bias, data extraction, and classification [127]. We tried to mitigate the risk for publication bias by selecting three bibliographic databases with different focuses; negative results for a WoS journal may have been particularly interesting for an IEEE journal (e.g. technical details into an authentication protocol). There is always a chance that we missed some papers due to our search protocol. The journal filter and search query were quite specific. Furthermore, we did not apply snowballing nor did we include grey literature. Nonetheless, many articles had to be discarded because they appeared off-topic when relying on the abstract. Selection bias was mitigated by discussing and testing the inclusion/exclusion criteria by the multidisciplinary set of authors on several occasions. 136 articles from three different bibliographic libraries remained for analysis and we expect that adding more articles would add little new information.

Threats to data extraction and classification are related to researcher bias. We mitigated these threats by thoroughly and repeatedly discussing the data extraction form and operationalization of each classification in the full author team. We decided to maintain the different levels of abstraction of the stakeholders, features, and obstacles from the articles. This way we limited uncertainty and bias in the classification, and give the reader better insight in the heterogeneity of the articles. Furthermore, we tried to mitigate the risk of overlooking a relevant text fragment or misinterpreting semantic relationships by a close reading of each article.

With the measures described, we believe we mitigated the main threats to validity for this review as much as possible.

3.5 Conclusion

In this study, we systematically reviewed 136 articles on HISs that appeared in three different bibliographic databases over the past twelve years. The current state of HISs was described by focusing on the elements: healthcare domains, stakeholders, features, and obstacles. With this SLR, we believe we present a broad and comprehensive overview of HISs.

According to the literature, most HISs are aimed at hospital care, whereas few studies referred to HISs in telehealth and telemedicine. Nonetheless, a wide range of direct system users, in-direct systems users, and technical stakeholders were identified. The literature described a wide variety of features related to general information systems, as well as a wide variety of features related to the health care setting. Despite dependency on HISs, more than 70 different obstacles could be distinguished. The huge heterogeneity in features and obstacles may not only stem from the diversity of health settings studied, but also from a lack of basic consensus.

In conclusion, the results of the presented work are useful for HIS researchers, users, developers, and scientists. The lists of stakeholders, features, and obstacles presented, may help managers in decision making and in system development or adaptation. The overview of features and obstacles may be explored further by scientists and system developers in order to design better HISs tailored to the healthcare domain. Scientists from different disciplines may use this study to identify new research areas on HISs and HISs data. In our future work, we will aim to develop an HIS reference architecture using the lessons learned from this SLR.

Appendix

Quality Assessment Scores

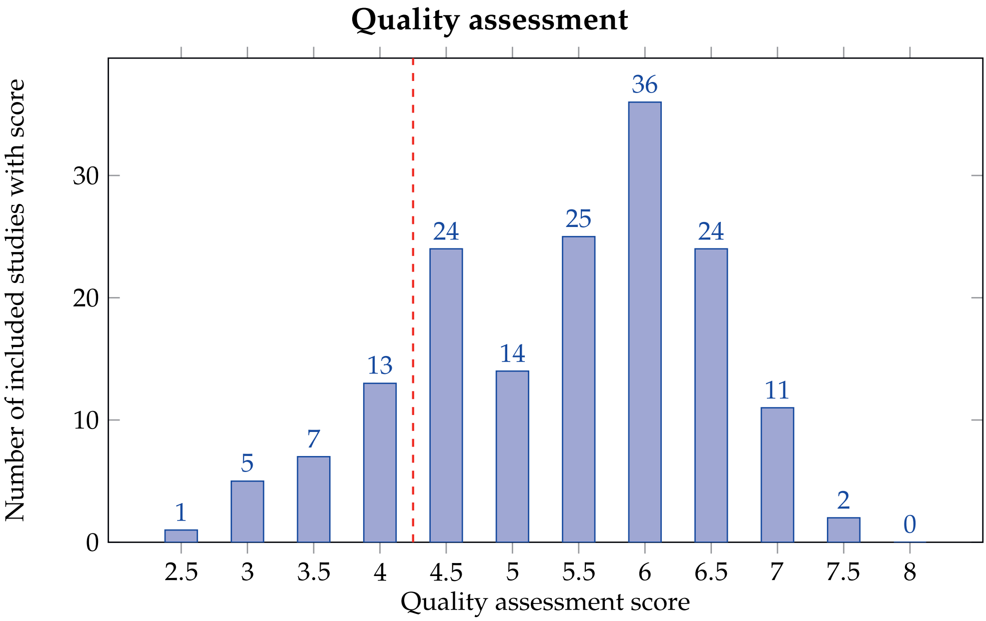


Figure 3.5: The quality score distribution for the 162 studies after applying the study selection criteria. The studies on the left of the red dashed line are excluded from the results due to their low quality.

Explanation of data extraction form

8: *Targeted domain*: The word(s) the authors used to describe the domain of the healthcare sector in which the discussed HIS is applied. Example: hospital, dentistry, pharmacy, etcetera
9: *Main theme of study*: Where is the study about? Example: Design of a reference framework. This is used for the in/exclusion of the studies.
10: *Assessment approach* : What kind of an assessment strategy is applied. This is used for the in/exclusion of the studies.
11: *Delivery model*: How is the software executed? Can be multiple answers. For example, server applications and web applications go very often together.
12: *Name of HIS*: The name of the HIS if applicable. Example: Healthvision
13: *Mentioned Stakeholders*: The word(s) the authors used to describe the stakeholders mentioned in the article. Example: Doctors, government, software developer etcetera.
14: *Constraints/limitations*: Are there any constraints or limitations to the article/HIS? Example: Software not running, Very simple structure, etcetera. This was used for the quality assessment
15: *Mentioned Obstacles*: The word(s) the authors used to describe the obstacles for HISs. Example: Data integration, no end-to-end solution available, Performance assurance, etcetera.
16: *Mentioned features*: The word(s) the authors used to describe the features of HISs mentioned in the article. For example: Financial management, Appointment scheduling, etcetera.
17: *Saved data*: The words the authors used to describe which data is saved in the HIS. For example: Blood values, Patient demographics, etcetera. (This category was in the synthesis merged with the features.)
18: *Used country/place*: For which country is the HISs developed? This field was later not used for this paper.

Analytic set

[1] Abugessaisa, I., Saevarsdottir, S., Tsipras, G., Lindblad, S., Sandin, C., Nikamo, P., Stahle, M., Malmstrom, V., Klareskog, L., and Tegner, J. (2014). Accelerating translational research by clinically driven development of an informatics platform–a case study. *PLoS One*, 9(9):e104382.

[2] Adler-Milstein, J., Everson, J., and Lee, S. Y. D. (2014). Sequencing of ehr adoption among us hospitals and the impact of meaningful use. *Journal of the American Medical Informatics Association*, 21(6):984–991.

[3] Adler-Milstein, J., Holmgren, A. J., Kralovec, P., Worzala, C., Searcy, T., and Patel, V. (2017). Electronic health record adoption in us hospitals: the emergence of a digital "advanced use" divide. *Journal of the American Medical Informatics Association*, 24(6):1142–1148.

[4] Ahmadian, L., Khajouei, R., Nejad, S. S., Ebrahimzadeh, M., and Nikkar, S. E. (2014). Prioritizing barriers to successful implementation of hospital information systems. *Journal of Medical Systems*, 38(12).

[5] Al-Rayes, S. A., Alumran, A., and AlFayez, W. (2019). The adoption of the electronic health record by physicians. *Methods of Information in Medicine*, 58(2-3):63–70.

[6] Alsaffar, M., Yellowlees, P., Odor, A., and Hogarth, M. (2017). The state of open source electronic health record projects: A software anthropology study. *Jmir Medical Informatics*, 5(1).

[7] Alsohime, F., Temsah, M. H., Al-Eyadhy, A., Bashiri, F. A., Househ, M., Jamal, A., Hasan, G., Alhaboob, A. A., Alabdulhafid, M., and Amer, Y. S. (2019). Satisfaction and perceived usefulness with newly-implemented electronic health records system among pediatricians at a university hospital. *Computer Methods and Programs in Biomedicine*, 169:51–57.

[8] Alyousef, B., Carayon, P., Hoonakker, P., Hundt, A. S., Salek, D., and Tomcavage, J. (2017). Obstacles experienced by care managers in managing information for the care of chronically ill patients. *International Journal of Human-Computer Interaction*, 33(4):313–321.

[9] Ancker, J. S., Kern, L. M., Edwards, A., Nosal, S., Stein, D. M., Hauser, D., Kaushal, R., and Investigators, H. (2014). How is the electronic health record being used? use of ehr data to assess physician-level variability in technology use. *Journal of the American Medical Informatics Association*, 21(6):1001–1008.

[10] Angoff, G. H., O'Connell, J. J., Gaeta, J. M., De las Nueces, D., Lawrence, M., Nembang, S., and Baggett, T. P. (2019). Electronic medical record implementation for a healthcare system caring for homeless people. *Jamia Open*, 2(1):89–98.

[11] Arnold, C. W., McNamara, M., El-Saden, S., Chen, S., Taira, R. K., and Bui, A. A. T. (2013). Imaging informatics for consumer health: towards a radiology patient portal. *Journal of the American Medical Informatics Association*, 20(6):1028–1036.

[12] Asare, E., Wang, L., and Fang, X. (2020). Conformance checking: Workflow of hospitals and workflow of open-source emrs. *IEEE Access*, 8:139546–139566.

[13] Atashi, A., Khajouei, R., Azizi, A., and Dadashi, A. (2016). User interface problems of a nation-wide inpatient information system: A heuristic evaluation. *Applied Clinical Informatics*, 7(1):89–100.

[14] Bagayoko, C. O., Tchuente, J., Traore, D., Lipenguet, G. M., Mbenga, R. O., Koumamba, A. P., Ondjani, M. C., Ndjeli, O. L., and Gagnon, M. P. (2020). Implementation of a national electronic health information system in gabon: a survey of healthcare providers' perceptions. *Bmc Medical Informatics and Decision Making*, 20(1).

[15] Baumgart, D. C. (2020). Digital advantage in the covid-19 response: perspective from canada's largest integrated digitalized healthcare system. *NPJ Digit Med*, 3.

[16] Berquedich, M., Berquedich, A., Kamach, O., Masmoudi, M., Chebbak, A., and Deshayes, L. (2020). Developing a mobile covid-19 prototype management application integrated with an electronic health record for effective management in hospitals. *IEEE Engineering Management Review*, 48(4):55–64.

[17] Bhatti, U. A., Huang, M. X., Wu, D., Zhang, Y., Mehmood, A., and Han, H. R. (2019). Recommendation system using feature extraction and pattern recognition in clinical care systems. *Enterprise Information Systems*, 13(3):329–351.

[18] Bjering, H., Ginige, A., Maeder, A., Bensoussan, A., Zhu, X., and Lattuca, C. (2011). Electronic medical record information system for patient consultations in chinese medicine. *Stud Health Technol Inform*, 168:10–5.

[19] Blank, A., Prytherch, H., Kaltschmidt, J., Krings, A., Sukums, F., Mensah, N., Zakane, A., Loukanova, S., Gustafsson, L. L., Sauerborn, R., and Haefeli, W. E. (2013). "quality of prenatal and maternal care: bridging the know-do gap" (qualmat study): an electronic clinical decision support system for rural sub-saharan africa. *Bmc Medical Informatics and Decision Making*, 13.

[20] Bondale, N., Kimbahune, S., and Pande, A. K. (2013). mhealthphc: An ict tool for primary healthcare in india. *IEEE Technology and Society Magazine*, 32(3):31–38.

[21] Callen, J., Li, L., Georgiou, A., Paoloni, R., Gibson, K., Li, J. L., Stewart, M., Braithwaite, J., and Westbrook, J. I. (2014). Does an integrated emergency department information system change the sequence of clinical work? a mixed-method cross-site study. *International Journal of Medical Informatics*, 83(12):958–966.

[22] Carroll, N. and Richardson, I. (2020). Enablers and barriers for hospital pharmacy information systems. *Health Informatics Journal*, 26(1):406–419.

[23] Cheung, C. S., Tong, E. L., Cheung, N. T., Chan, W. M., Wang, H. H., Kwan, M. W., Fan, C. K., Liu, K. Q., and Wong, M. C. (2013). Factors associated with adoption of the electronic health record system among primary care physicians. *JMIR Med Inform*, 1(1):e1.

[24] Cifuentes, M., Davis, M., Fernald, D., Gunn, R., Dickinson, P., and Cohen, D. J. (2015). Electronic health record challenges, workarounds, and solutions observed in practices integrating behavioral health and primary care. *J Am Board Fam Med*, 28 Suppl 1:S63–72.

[25] Clark, K. D., Woodson, T. T., Holden, R. J., Gunn, R., and Cohen, D. J. (2019). Translating research into agile development (triad): Development of electronic health record tools for primary care settings. *Methods of Information in Medicine*, 58(1):1–8.

[26] Colicchio, T. K. and Cimino, J. J. (2020). Twilighted homegrown systems: The experience of six traditional electronic health record developers in the post-meaningful use era. *Applied Clinical Informatics*, 11(2):356–365.

[27] Denecke, K., Dittli, P. A., Kanagarasa, N., and Nussli, S. (2018). Facilitating the information exchange using a modular electronic discharge summary. *Stud Health Technol Inform*, 248:72–79.

[28] deRiel, E., Puttkammer, N., Hyppolite, N., Diallo, J., Wagner, S., Honore, J. G., Balan, J. G., Celestin, N., Valles, J. S., Duval, N., Thimothe, G., Boncy, J., Coq, N. R. L., and Barnhart, S. (2018). Success factors for implementing and sustaining a mature electronic medical record in a low-resource setting: a case study of isante in haiti. *Health Policy Plan*, 33(2):237–246.

[29] Documet, J., Le, A., Liu, B., Chiu, J., and Huang, H. K. (2010). A multimedia electronic patient record (epr) system for image-assisted minimally invasive spinal surgery. *Int J Comput Assist Radiol Surg*, 5(3):195–209.

[30] Duke, J. D., Morea, J., Mamlin, B., Martin, D. K., Simonaitis, L., Takesue, B. Y., Dixon, B. E., and Dexter, P. R. (2014). Regenstrief institute's medical gopher: A next-generation homegrown electronic medical record system. *International Journal of Medical Informatics*, 83(3):170–179.

[31] Dunsmuir, D. T., Payne, B. A., Cloete, G., Petersen, C. L., Görges, M., Lim, J., Dadelszen, P. v., Dumont, G. A., and Ansermino, J. M. (2014). Development of mhealth applications for pre-eclampsia triage. *IEEE Journal of Biomedical and Health Informatics*, 18(6):1857–1864.

[32] El-Sappagh, S., Alonso, J. M., Ali, F., Ali, A., Jang, J., and Kwak, K. (2018). An ontology-based interpretable fuzzy decision support system for diabetes diagnosis. *IEEE Access*, 6:37371–37394.

[33] Emberti Gialloreti, L., Basa, F. B., Moramarco, S., Salih, A. O., Alsilefanee, H. H., Qadir, S. A., Bezenchek, A., Incardona, F., Di Giovanni, D., Khorany, R., Alhanabadi, L. H. H., Salih, S. O., Akhshirsh, G. S., Azeez, B. S., Tofiq, B. A., and Palombi, L. (2020). Supporting iraqi kurdistan health authorities in post-conflict recovery: The development of a health monitoring system. *Front Public Health*, 8.

[34] Faujdar, D. S., Sahay, S., Singh, T., Jindal, H., and Kumar, R. (2019). Public health information systems for primary health care in india: A situational analysis study. *J Family Med Prim Care*, 8(11):3640–6.

[35] Faujdar, D. S., Sahay, S., Singh, T., Kaur, M., and Kumar, R. (2020). Field testing of a digital health information system for primary health care: A quasi-experimental study from india. *International Journal of Medical Informatics*, 141.

[36] Featherstone, I. and Keen, J. (2012). Do integrated record systems lead to integrated services? an observational study of a multi-professional system in a diabetes service. *International Journal of Medical Informatics*, 81(1):45–52.

[37] Fernandez, A. D., Fernandez, D. R., Marcos-Jorquera, D., and Iglesias, V. G. (2020). Support system for early diagnosis of chronic obstructive pulmonary disease based on the service-oriented architecture paradigm and business process management strategy: Development and usability survey among patients and health care providers. *Journal of Medical Internet Research*, 22(3).

[38] Flohr, L., Beaudry, S., Johnson, K. T., West, N., Burns, C. M., Ansermino, J. M., Dumont, G. A., Wensley, D., Skippen, P., and Görges, M. (2018). Clinician-driven design of<italic>vitalpad</italic>–an intelligent monitoring and communication device to improve patient safety in the intensive care unit. *IEEE Journal of Translational Engineering in Health and Medicine*, 6:1–14.

[39] Fortney, J. C., Pyne, J. M., Steven, C. A., Williams, J. S., Hedrick, R. G., Lunsford, A. K., Raney, W. N., Ackerman, B. A., Ducker, L. O., Bonner, L. M., and Smith, J. L. (2010). A web-based clinical decision support system for depression care management. *Am J Manag Care*, 16(11):849–54.

[40] Frame, A., LaMantia, M., Reddy Bynagari, B. B., Dexter, P., and Boustani, M. (2013). Development and implementation of an electronic decision support to manage the health of a high-risk population: The enhanced electronic medical record aging brain care software (emr-abc). *EGEMS (Wash DC)*, 1(1):1009.

[41] Fraser, H. S. F., Thomas, D., Tomaylla, J., Garcia, N., Lecca, L., Murray, M., and Becerra, M. C. (2012). Adaptation of a web-based, open source electronic medical record system platform to support a large study of tuberculosis epidemiology. *Bmc Medical Informatics and Decision Making*, 12.

[42] Friedman, A., Crosson, J. C., Howard, J., Clark, E. C., Pellerano, M., Karsh, B. T., Crabtree, B., Jaen, C. R., and Cohen, D. J. (2014). A typology of electronic health record workarounds in small-to-medium size primary care practices. *Journal of the American Medical Informatics Association*, 21(E1):E78–E83.

[43] Garcia, P. J., Vargas, J. H., Caballero, P., Calle, J., and Bayer, A. M. (2009). An e-health driven laboratory information system to support hiv treatment in peru: E-quity for laboratory personnel, health providers and people living with hiv. *Bmc Medical Informatics and Decision Making*, 9.

[44] Goh, W. P., Tao, X. H., Zhang, J., and Yong, J. M. (2016). Decision support systems for adoption in dental clinics: A survey. *Knowledge-Based Systems*, 104:195–206.

[45] Groenhof, T. K. J., Rittersma, Z. H., Bots, M. L., Brandjes, M., Jacobs, J. J. L., Grobbee, D. E., van Solinge, W. W., Visseren, F. L. J., Haitjema, S., and Asselbergs, F. W. (2019). A computerised decision support system for cardiovascular risk management 'live' in the electronic health record environment: development, validation and implementation—the utrecht cardiovascular cohort initiative. *Neth Heart J*, 27(9):435–42.

[46] Grout, R. W., Cheng, E. R., Carroll, A. E., Bauer, N. S., and Downs, S. M. (2018). A six-year repeated evaluation of computerized clinical decision support system user acceptability. *International Journal of Medical Informatics*, 112:74–81.

[47] Guo, J., Iribarren, S., Kapsandoy, S., Perri, S., and Staggers, N. (2011). Usability evaluation of an electronic medication administration record (emar) application. *Applied Clinical Informatics*, 2(2):202–224.

[48] Görs, M., Albert, M., Schwedhelm, K., Herrmann, C., and Schilling, K. (2016). Design of an advanced telemedicine system for remote supervision. *IEEE Systems Journal*, 10(3):1089–1097.

[49] Han, C., Rundo, L., Murao, K., Nemoto, T., and Nakayama, H. (2020). Bridging the gap between ai and healthcare sides: Towards developing clinically relevant ai-powered diagnosis systems. *Artificial Intelligence Applications and Innovations*, 584:320–33.

[50] Heidebrecht, C. L., Kwong, J. C., Finkelstein, M., Quan, S. D., Pereira, J. A., Quach, S., and Deeks, S. L. (2014). Electronic immunization data collection systems: application of an evaluation framework. *Bmc Medical Informatics and Decision Making*, 14.

[51] Hester, G., Lang, T., Madsen, L., Tambyraja, R., and Zenker, P. (2019). Timely data for targeted quality improvement interventions: Use of a visual analytics dashboard for bronchiolitis. *Applied Clinical Informatics*, 10(1):168–174.

[52] Hochgesang, M., Zamudio-Haas, S., Moran, L., Nhampossa, L., Packel, L., Leslie, H., Richards, J., and Shade, S. B. (2017). Scaling-up health information systems to improve hiv treatment: An assessment of initial patient monitoring systems in mozambique. *International Journal of Medical Informatics*, 97:322–330.

[53] Holanda, A. A., Sa, H., Vieira, A., and Catrib, A. M. F. (2012). Use and satisfaction with electronic health record by primary care physicians in a health district in brazil. *Journal of Medical Systems*, 36(5):3141–3149.

[54] Hsiao, C. J., Marsteller, J. A., and Simon, A. E. (2014). Electronic medical record features and seven quality of care measures in physician offices. *Am J Med Qual*, 29(1):44–52.

[55] Hum, R. S., Cato, K., Sheehan, B., Patel, S., Duchon, J., DeLaMora, P., Ferng, Y. H., Graham, P., Vawdrey, D. K., Perlman, J., Larson, E., and Saiman, L. (2014). Developing clinical decision support within a commercial electronic health record system to improve antimicrobial prescribing in the neonatal icu. *Appl Clin Inform*, 5(2):368–87.

[56] Jacobs, M., Boersma, L. J., Swart, R., Mannens, R., Reymen, B., Korver, F., van Merode, F., and Dekker, A. (2019). Electronic health record implementation in a large academic radiotherapy department: Temporarily disruptions but long-term benefits. *International Journal of Medical Informatics*, 129:342–348.

[57] Jahn, F., Issler, L., Winter, A., and Takabayashi, K. (2009). Comparing a japanese and a german hospital information system assessing structural quality by 3lgm(2) models. *Methods of Information in Medicine*, 48(6):531–539.

[58] Jian, W. S., Wen, H. C., Scholl, J., Shabbir, S. A., Lee, P., Hsu, C. Y., and Li, Y. C. (2011). The taiwanese method for providing patients data from multiple hospital ehr systems. *Journal of Biomedical Informatics*, 44(2):326–332.

[59] Kabukye, J. K., de Keizer, N., and Cornet, R. (2020). Elicitation and prioritization of requirements for electronic health records for oncology in low resource settings: A concept mapping study. *International Journal of Medical Informatics*, 135.

[60] Kahouei, M., Zadeh, J. M., and Roghani, P. S. (2015). The evaluation of the compatibility of electronic patient record (epr) system with nurses' management needs in a developing country. *International Journal of Medical Informatics*, 84(4):263–270.

[61] Kaipio, J., Laaveri, T., Hypponen, H., Vainiomaki, S., Reponen, J., Kushniruk, A., Borycki, E., and Vanska, J. (2017). Usability problems do not heal by themselves: National survey on physicians' experiences with ehrs in finland. *International Journal of Medical Informatics*, 97:266–281.

[62] Kallen, M. A., Yang, D., and Haas, N. (2012). A technical solution to improving palliative and hospice care. *Support Care Cancer*, 20(1):167–74.

[63] Katzan, I. L., Fan, Y., Speck, M., Morton, J., Fromwiller, L., Urchek, J., Uchino, K., Griffith, S. D., and Modic, M. (2015). Electronic stroke carepath: Integrated approach to stroke care. *Circ Cardiovasc Qual Outcomes*, 8(6 Suppl 3):S179–89.

[64] Khajouei, R., Gohari, S. H., and Mirzaee, M. (2018). Comparison of two heuristic evaluation methods for evaluating the usability of health information systems. *Journal of Biomedical Informatics*, 80:37–42.

[65] Kihuba, E., Gathara, D., Mwinga, S., Mulaku, M., Kosgei, R., Mogoa, W., Nyamai, R., and English, M. (2014). Assessing the ability of health information systems in hospitals to support evidence-informed decisions in kenya. *Glob Health Action*, 7:24859.

[66] Kim, S., Lee, K. H., Hwang, H., and Yoo, S. (2016). Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (emr) using the unified theory of acceptance and use of technology (utaut) in a tertiary hospital. *Bmc Medical Informatics and Decision Making*, 16.

[67] Kropf, S., Chalopin, C., Lindner, D., and Denecke, K. (2017). Domain modeling and application development of an archetype- and xml-based ehrs. practical experiences and lessons learnt. *Appl Clin Inform*, 8(2):660–679.

[68] Kulkarni, P. and Ozturk, Y. (2011). mphasis: Mobile patient healthcare and sensor information system. *Journal of Network and Computer Applications*, 34(1):402–417.

[69] Laing, G. L., Bruce, J. L., Skinner, D. L., Allorto, N. L., Clarke, D. L., and Aldous, C. (2014). Development, implementation, and evaluation of a hybrid electronic medical record system specifically designed for a developing world surgical service. *World J Surg*, 38(6):1388–97.

[70] Lee, K., Jung, S. Y., Hwang, H., Yoo, S., Baek, H. Y., Baek, R. M., and Kim, S. (2017). A novel concept for integrating and delivering health information using a comprehensive digital dashboard: An analysis of healthcare professionals' intention to adopt a new system and the trend of its real usage. *International Journal of Medical Informatics*, 97:98–108.

[71] Lei, J. B., Sockolow, P., Guan, P. C., Meng, Q., and Zhang, J. J. (2013). A comparison of electronic health records at two major peking university hospitals in china to united states meaningful use objectives. *Bmc Medical Informatics and Decision Making*, 13.

[72] Longhurst, C., Turner, S., and Burgos, A. E. (2009). Development of a web-based decision support tool to increase use of neonatal hyperbilirubinemia guidelines. *Jt Comm J Qual Patient Saf*, 35(5):256–62.

[73] Ma, K. C., Fernandez, J. R., Amezcua, L., Lerner, A., Shiroishi, M. S., and Liu, B. J. (2015). Design and development of an ethnically-diverse imaging informatics-based efolder system for multiple sclerosis patients. *Comput Med Imaging Graph*, 46 Pt 2:257–68.

[74] Magalhaes, D., Martins, J., Branco, F., Au-Yong-Oliveira, M., Goncalves, R., and Moreira, F. (2020). A proposal for a 360 degrees information system model for private health care organizations. *Expert Systems*, 37(5).

[75] Makam, A. N., Lanham, H. J., Batchelor, K., Samal, L., Moran, B., Howell-Stampley, T., Kirk, L., Cherukuri, M., Santini, N., Leykum, L. K., and Halm, E. A. (2013). Use and satisfaction with key functions of a common commercial electronic health record: a survey of primary care providers. *Bmc Medical Informatics and Decision Making*, 13.

[76] Marca, G., Perez, A. J., Blanco-Garcia, M. G., Miravalles, E., Soley, P., and Ortiga, B. (2014). The use of electronic health records in spanish hospitals. *Health Information Management Journal*, 43(3):37–44.

[77] Martinez, W., Threatt, A. L., Rosenbloom, S. T., Wallston, K. A., Hickson, G. B., and Elasy, T. A. (2018). A patient-facing diabetes dashboard embedded in a patient web portal: Design sprint and usability testing. *JMIR Hum Factors*, 5(3):e26.

[78] McAlearney, A. S., Robbins, J., Hirsch, A., Jorina, M., and Harrop, J. P. (2010). Perceived efficiency impacts following electronic health record implementation: An exploratory study of an urban community health center network. *International Journal of Medical Informatics*, 79(12):807–816.

[79] McCoy, A. B., Wright, A., and Sittig, D. F. (2015). Cross-vendor evaluation of key user-defined clinical decision support capabilities: a scenario-based assessment of certified electronic health records with guidelines for future development. *Journal of the American Medical Informatics Association*, 22(5):1081–1088.

[80] McQuaida, L., Breen, P., Grimson, J., Normand, C., Dunne, M., Delanty, N., Kalra, D., and Fitzsimons, M. (2010). Socio-technical considerations in epilepsy electronic patient record implementation. *International Journal of Medical Informatics*, 79(5):349–360.

[81] Montazeri, M., Khajouei, R., and Montazeri, M. (2020). Evaluating hospital information system according to iso 9241 part 12. *Digital Health*, 6.

[82] Monteiro, E. J. M., Costa, C., and Oliveira, J. L. (2016). A cloud architecture for teleradiology-as-a-service. *Methods of Information in Medicine*, 55(3):203–214.

[83] Muinga, N., Magare, S., Monda, J., English, M., Fraser, H., Powell, J., and Paton, C. (2020). Digital health systems in kenyan public hospitals: a mixed-methods survey. *Bmc Medical Informatics and Decision Making*, 20(1).

[84] Nabovati, E., Vakili-Arki, H., Eslami, S., and Khajouei, R. (2014). Usability evaluation of laboratory and radiology information systems integrated into a hospital information system. *Journal of Medical Systems*, 38(4).

[85] Nguefack-Tsague, G., Tamfon, B. B., Ngnie-Teta, I., Ngoufack, M. N., Keugoung, B., Bataliack, S. M., and Ndongo, C. B. (2020). Factors associated with the performance of routine health information system in yaounde-cameroon: a cross-sectional survey. *Bmc Medical Informatics and Decision Making*, 20(1).

[86] Norris, K. C., Duru, O. K., Alicic, R. Z., Daratha, K. B., Nicholas, S. B., McPherson, S. M., Bell, D. S., Shen, J. I., Jones, C. R., Moin, T., Waterman, A. D., Neumiller, J. J., Vargas, R. B., Bui, A. A. T., Mangione, C. M., and Tuttle, K. R. (2019). Rationale and design of a multicenter chronic kidney disease (ckd) and at-risk for ckd electronic health records-based registry: Cure-ckd. *BMC Nephrol*, 20.

[87] O'Malley, A. S., Draper, K., Gourevitch, R., Cross, D. A., and Scholle, S. H. (2015). Electronic health records and support for primary care teamwork. *Journal of the American Medical Informatics Association*, 22(2):426–434.

[88] Palojoki, S., Makela, M., Lehtonen, L., and Saranto, K. (2017). An analysis of electronic health record-related patient safety incidents. *Health Informatics Journal*, 23(2):134–145.

[89] Pare, G., Raymond, L., de Guinea, A. O., Poba-Nzaou, P., Trudel, M. C., Marsan, J., and Micheneau, T. (2015). Electronic health record usage behaviors in primary care medical practices: A survey of family physicians in canada. *International Journal of Medical Informatics*, 84(10):857–867.

[90] Park, Y. T. and Han, D. (2017). Current status of electronic medical record systems in hospitals and clinics in korea. *Healthcare Informatics Research*, 23(3):189–198.

[91] Pfoh, E. R., Abramson, E., Zandieh, S., Edwards, A., and Kaushal, R. (2012). Satisfaction after the transition between electronic health record systems at six ambulatory practices. *Journal of Evaluation in Clinical Practice*, 18(6):1133–1139.

[92] Pickering, B. W., Dong, Y., Ahmed, A., Giri, J., Kilickaya, O., Gupta, A., Gajic, O., and Herasevich, V. (2015). The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: A pilot step-wedge cluster randomized trial. *International Journal of Medical Informatics*, 84(5):299–307.

[93] Poba-Nzaou, P., Kume, N., and Kobayashi, S. (2020). Developing and sustaining an open source electronic health record: Evidence from a field study in japan. *Journal of Medical Systems*, 44(9).

[94] Poba-Nzaou, P. and Uwizeyemungu, S. (2019). Variation in electronic health record adoption in european public hospitals: a configurational analysis of key functionalities. *Health and Technology*, 9(4):439–448.

[95] Protti, D., Johansen, I., and Perez-Torres, F. (2009). Comparing the application of health information technology in primary care in denmark and andalucia, spain. *International Journal of Medical Informatics*, 78(4):270–283.

[96] Rajkovic, P., Jankovic, D., and Milenkovic, A. (2013). Developing and deploying medical information systems for serbian public healthcare - challenges, lessons learned and guidelines. *Computer Science and Information Systems*, 10(3):1429–1454.

[97] Raymond, L., Pare, G., de Guinea, A. O., Poba-Nzaou, P., Trudel, M. C., Marsan, J., and Micheneau, T. (2015). Improving performance in medical practices through the extended use of electronic medical record systems: a survey of canadian family physicians. *Bmc Medical Informatics and Decision Making*, 15.

[98] Raymond, L., Pare, G., and Maillet, E. (2020). Enabling laboratory medicine in primary care through it systems use. *Data Base for Advances in Information Systems*, 51(1):70–83.

[99] Raymond, L., Pare, G., and Marchand, M. (2019). Extended use of electronic health records by primary care physicians: Does the electronic health record artefact matter? *Health Informatics Journal*, 25(1):71–82.

[100] Resnick, H. E. and Alwan, M. (2010). Use of health information technology in home health and hospice agencies: United states, 2007. *Journal of the American Medical Informatics Association*, 17(4):389–395.

[101] Riahi, S., Fischler, I., Stuckey, M. I., Klassen, P. E., and Chen, J. (2017). The value of electronic medical record implementation in mental health care: A case study. *Jmir Medical Informatics*, 5(1).

[102] Ritchie, J. and Welch, B. (2020). Categorization of third-party apps in electronic health record app marketplaces: Systematic search and analysis. *Jmir Medical Informatics*, 8(5).

[103] Rogers, M. L., Sockolow, P. S., Bowles, K. H., Hand, K. E., and George, J. (2013). Use of a human factors approach to uncover informatics needs of nurses in documentation of care. *International Journal of Medical Informatics*, 82(11):1068–1074.

[104] Sarwar, M. A., Bashir, T., Shahzad, O., and Abbas, A. (2019). Cloud-based architecture to implement electronic health record (ehr) system in pakistan. *IT Professional*, 21(3):49–54.

[105] Schreiweis, B., Trinczek, B., Kopcke, F., Leusch, T., Majeed, R. W., Wenk, J., Bergh, B., Ohmann, C., Rohrig, R., Dugas, M., and Prokosch, H. U. (2014). Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials. *International Journal of Medical Informatics*, 83(11):860–868.

[106] Senathirajah, Y., Kaufman, D., and Bakken, S. (2016). User-composable electronic health record improves efficiency of clinician data viewing for patient case appraisal: A mixed-methods study. *EGEMS (Wash DC)*, 4(1):1176.

[107] Shahnaz, A., Qamar, U., and Khalid, A. (2019). Using blockchain for electronic health records. *IEEE Access*, 7:147782–147795.

[108] Sidek, Y. H. and Martins, J. T. (2017). Perceived critical success factors of electronic health record system implementation in a dental clinic context: An organisational management perspective. *International Journal of Medical Informatics*, 107:88–100.

[109] Silva, B. S., de Azevedo Guimarães, E. A., de Oliveira, V. C., Cavalcante, R. B., Pinheiro, M. M. K., Gontijo, T. L., Rodrigues, S. B., Ferreira, A. P., de Oliveira Quites, H. F., and Pinto, I. C. (2020). National immunization program information system: implementation context assessment. *BMC Health Serv Res*, 20.

[110] Sinard, J. H., Castellani, W. J., Wilkerson, M. L., and Henricks, W. H. (2015). Stand-alone laboratory information systems versus laboratory modules incorporated in the electronic health record. *Arch Pathol Lab Med*, 139(3):311–8.

[111] Sockolow, P. S., Rogers, M., Bowles, K. H., Hand, K. E., and George, J. (2014). Challenges and facilitators to nurse use of a guideline-based nursing information system: recommendations for nurse executives. *Appl Nurs Res*, 27(1):25–32.

[112] Sockolow, P. S., Weiner, J. P., Bowles, K. H., Abbott, P., and Lehmann, H. P. (2011). Advice for decision makers based on an electronic health record evaluation at a program for all-inclusive care for elders site. *Applied Clinical Informatics*, 2(1):18–38.

[113] Soderberg, K., Rajamani, S., Wholey, D., and LaVenture, M. (2016). Health reform in minnesota: An analysis of complementary initiatives implementing electronic health record technology and care coordination. *Online J Public Health Inform*, 8(3):e204.

[114] Tamfon, B. B., Ndongo, C. B., Bataliack, S. M., Ngoufack, M. N., and Nguefack-Tsague, G. (2020). Routine health information system in the health facilities in yaounde-cameroon: assessing the gaps for strengthening. *Bmc Medical Informatics and Decision Making*, 20(1).

[115] Tan, B. T., Fralick, J., Flores, W., Schrandt, C., Davis, V., Bruynell, T., Wilson, L., Christopher, J., Weber, S., and Shah, N. (2017). Implementation of epic beaker clinical pathology at stanford university medical center. *Am J Clin Pathol*, 147(3):261–272.

[116] Tierney, W. M., Sidle, J. E., Diero, L. O., Sudoi, A., Kiplagat, J., Macharia, S., Shen, C., Yeung, A., Were, M. C., Slaven, J. E., and Wools-Kaloustian, K. (2016). Assessing the impact of a primary care electronic medical record system in three kenyan rural health centers. *Journal of the American Medical Informatics Association*, 23(3):544–552.

[117] Topaz, M., Ronquillo, C., Peltonen, L. M., Pruinelli, L., Sarmiento, R. F., Badger, M. K., Ali, S., Lewis, A., Georgsson, M., Jeon, E., Tayaben, J. L., Kuo, C. H., Islam, T., Sommer, J., Jung, H., Eler, G. J., Alhuwail, D., and Lee, Y. L. (2016). Nurse informaticians report low satisfaction and

multi-level concerns with electronic health records: Results from an international survey. *AMIA Annu Symp Proc*, 2016:2016–2025.

[118] Triantafyllidis, A. K., Koutkias, V. G., Chouvarda, I., and Maglaveras, N. (2013). A pervasive health system integrating patient monitoring, status logging, and social sharing. *IEEE Journal of Biomedical and Health Informatics*, 17(1):30–37.

[119] Trudel, M. C., Marsan, J., Pare, G., Raymond, L., de Guinea, A. O., Maillet, E., and Micheneau, T. (2017). Ceiling effect in emr system assimilation: a multiple case study in primary care family practices. *Bmc Medical Informatics and Decision Making*, 17.

[120] Tsavatewa, C., Musa, P. F., and Ramsingh, I. (2012). Integration of footprints information systems in palliative care: The case of medical center of central georgia. *Journal of Medical Systems*, 36(3):1511–1521.

[121] Tschanz, M., Dorner, T. L., Holm, J., and Denecke, K. (2018). Using emma to manage medication. *Computer*, 51(8):18–25.

[122] Tubaishat, A. and Al-Rawajfah, O. M. (2017). The use of electronic medical records in jordanian hospitals a nationwide survey. *Cin-Computers Informatics Nursing*, 35(10):538–545.

[123] Varshney, U. (2013). A smart approach to medication management. *Computer*, 46(1):71–76.

[124] Velickovski, F., Ceccaroni, L., Roca, J., Burgos, F., Galdiz, J. B., Marina, N., and Lluch-Ariet, M. (2014). Clinical decision support systems (cdss) for preventive management of copd patients. *J Transl Med*, 12 Suppl 2:S9.

[125] Viitanen, J., Hypponen, H., Laaveri, T., Vanska, J., Reponen, J., and Winblad, I. (2011). National questionnaire study on clinical ict systems proofs: Physicians suffer from poor usability. *International Journal of Medical Informatics*, 80(10):708–725.

[126] Vossebeld, D. M., Puik, E. C. N., Jaspers, J. E. N., and Schuurmans, M. J. (2019). Development process of a mobile electronic medical record for nurses: a single case study. *Bmc Medical Informatics and Decision Making*, 19.

[127] Walji, M. F., Kalenderian, E., Tran, D., Kookal, K. K., Nguyen, V., Tokede, O., White, J. M., Vaderhobli, R., Ramoni, R., Starke, P. C., Kimmes, N. S., Schoonheim-Klein, M. E., and Patel, V. L. (2013). Detection and characterization of usability problems in structured data entry interfaces in dentistry. *International Journal of Medical Informatics*, 82(2):128–138.

[128] Weemaes, M., Martens, S., Cuypers, L., Van Elslande, J., Hoet, K., Welkenhuysen, J., Goossens, R., Wouters, S., Houben, E., Jeuris, K., Laenen, L., Bruyninckx, K., Beuselinck, K., André, E., Depypere, M., Desmet, S., Lagrou, K., Van Ranst, M., Verdonck, A., and Goveia, J. (2020). Laboratory information system requirements to manage the covid-19 pandemic: A report from the belgian national reference testing center. *J Am Med Inform Assoc*, 27(8):1293–9.

[129] Whitt, K. J., Eden, L., Merrill, K. C., and Hughes, M. (2017). Nursing student experiences regarding safe use of electronic health records a pilot study of the safety and assurance factors for ehr resilience guides. *Cin-Computers Informatics Nursing*, 35(1):45–53.

[130] Williams, M. S., Kern, M. S., Lerch, V. R., Billet, J., Williams, J. L., and Moore, G. J. (2018). Implementation of a patient-facing genomic test report in the electronic health record using a web-application interface. *Bmc Medical Informatics and Decision Making*, 18.

[131] Wright, A., McCoy, A. B., Hickman, T. T. T., St Hilaire, D., Borbolla, D., Bowes, W. A., Dixon, W. G., Dorr, D. A., Krall, M., Malholtra, S., Bates, D. W., and Sittig, D. F. (2015). Problem list completeness in electronic health records: A multi-site study and assessment of success factors. *International Journal of Medical Informatics*, 84(10):784–790.

[132] Wright, A., Sittig, D. F., Ash, J. S., Feblowitz, J., Meltzer, S., McMullen, C., Guappone, K., Carpenter, J., Richardson, J., Simonaitis, L., Evans, R. S., Nichol, W. P., and Middleton, B. (2011). Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems. *Journal of the American Medical Informatics Association*, 18(3):232–242.

[133] Wright, A., Sittig, D. F., Ash, J. S., Sharma, S., Pang, J. E., and Middleton, B. (2009). Clinical decision support capabilities of commercially-available clinical information systems. *Journal of the American Medical Informatics Association*, 16(5):637–644.

[134] Yang, T. H., Sun, Y. L. S., and Lai, F. P. (2011). A scalable healthcare information system based on a service-oriented architecture. *Journal of Medical Systems*, 35(3):391–407.

[135] Yin, H. and Jha, N. K. (2017). A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles. *IEEE Transactions on Multi-Scale Computing Systems*, 3(4):228–241.

[136] Yusof, M. M. (2015). A case study evaluation of a critical care information system adoption using the socio-technical and fit approach. *International Journal of Medical Informatics*, 84(7):486–499.

ARCHITECTURE DESIGN

4

Study 3: Reference architecture design for health information systems

Background: Healthcare relies on Health Information Systems (HISs) to support the care and receive reimbursement for the care provided. Healthcare providers experience many problems with their HISs due to improper architecture design. To support the design of a proper HIS architecture, a Reference Architecture (RA) can be used that meets the various stakeholder concerns of HISs. Therefore, the objective of this study is to develop and analyze an RA following well-established architecture design methods.
Methods: Domain analysis was performed to scope and model the domain of HISs. For the architecture design, we applied the Views & Beyond approach and designed the RA's views based on the stakeholders and features from the domain analysis. We evaluated the RA with a case study.
Results: We derived the following four architecture views for HISs: The context diagram, decomposition view, layered view, and deployment view. Each view shows the architecture of the HIS from a different angle, suitable for various stakeholders. Based on a Japanese hospital information system study, we applied the RA and derived the application architecture.
Conclusion: We demonstrated that the methods of the software architecture design community could be used in the healthcare domain effectively and showed the applicability of the RA.

This chapter is based on an article published in an international journal:

Tummers, J., Tobi, H., Catal, C., & Tekinerdogan, B. (2021). Designing a reference architecture for health information systems. BMC Medical Informatics and Decision Making, 21(1), 1-14.

4.1 Background

Healthcare relies on Health Information Systems (HISs) to support various care processes and receive reimbursement for the care provided. Examples of functionalities are financial management, daily reporting, and medication management [45, 69, 128]. Unfortunately, current HISs still have some drawbacks. For example, lack of interoperability resulting in care professionals having difficulty communicating files [88, 129]. Other studies on HISs reported problems with poor interface design [59, 130], poor security [131, 132], missing features [66, 133], lack of professional support [47, 60], limited use [130, 134], and low data quality [69, 135]. Most of these problems occur when relevant standards, procedures, and guidelines are not followed effectively.

Because HISs consist of many interrelated software modules that should communicate, coordinate, and evolve over time [136], the software architecture is critical in HIS design. Bass et al. [33] define the software architecture of a program or a computing system as: ”The structure of the system, which comprises software elements, the externally visible properties of those elements, and the relationships among them.” The software architecture supports communication on the system, guides design decisions, informs maintenance, and facilitates architectural analysis of the overall system [137]. There are two main approaches for software architecture design: informal and formal. The back-draw of informal software architecture design relying on boxes-and-lines models, is that such a representation of the system is hard to understand because it is not standardized and does not follow a particular language. The formal approach follows the well-established ISO/ISEC/IEEE 42010 standard [138], which ensures unambiguous communication.

A particular type of architecture that is generic and can help design specific software architectures for multiple software systems is the Reference Architecture (RA). An RA is a generic design that can be used to derive specific Application Architecture (AAs) based on the identified stakeholders’ concerns, more quickly and with higher quality [78, 139]. The RA serves as an architecture blueprint for future software architects and should provide a standardized lexicon, taxonomy, and (architectural) vision [78, 140] . In the (grey) literature, several RA designs have been proposed for HISs [141–148]. More information on these RAs is available in the Related work Section.

In practice, the derivation of the AAs from RAs is not trivial for two basic reasons. First of all, some of the proposed RAs do not focus on HIS in general, but only address the hospital sub-domain [147, 149]. Secondly, the proposed RAs do not seem to follow a proper architecture documentation guideline. [144–146, 148]. Furthermore, these RAs are far from complete, which hampers the design of the required AAs.

The problems stakeholders experience with HISs require more clarity in healthcare's complex digital landscape, a clarity that RA provides. Therefore, the objective of this article is to develop an RA for HISs following well-established architecture design methods. The RA is dedicated to the healthcare domain and is represented using the software architecture viewpoints. To illustrate and evaluate the RA, an AA was derived in a case study on a Japanese hospital. The paper concludes with lessons learned and a discussion of the proposed RA.

4.2 Methods

4.2.1 Research questions

The following research questions were identified:

- RQ1: What are the stakeholders and their concerns related to HISs?
- RQ2: What is a feasible Reference Architecture for HISs?
- RQ3: Does the Reference Architecture allow for the derivation of a specific Application Architecture?

Our approach to these questions is depicted in Figure 4.1. Domain analysis is defined as the systematic activity for deriving and storing domain knowledge to support the engineering design process [150]. Domain analysis consists of domain scoping and domain modeling. Domain scoping identifies the domain's scope and the necessary knowledge sources to derive the key concepts [151, 152]. Domain modeling aims at representing the domain knowledge in a reusable format.

Based on the domain analysis, we choose the relevant viewpoints [153] for our architecture design step. We continued with a case study, to evaluate the RA's suitability for deriving an AA.

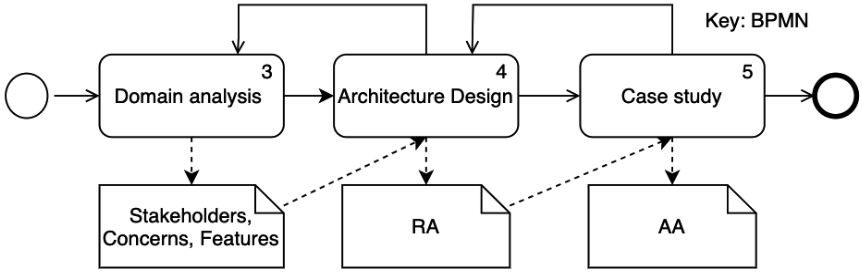


Figure 4.1: The adopted approach for the RA design. Numbers inside tasks represent corresponding Section numbers.

4.2.2 Method for deriving and evaluating application architecture

The RA can be used as a starting point for creating an AA [139]. The AA is described in this study as the software model of a specific application displayed through a combination of architectural views. To begin, an RA was created.

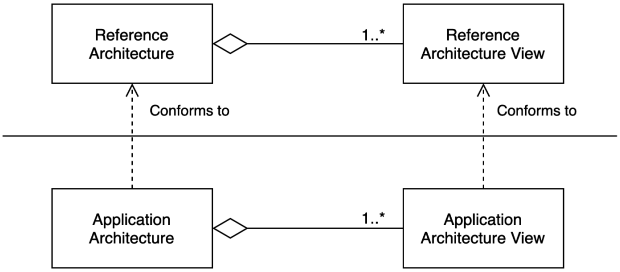


Figure 4.2: Methods used for deriving the AA. Each view from the RA will lead to a view in the AA, adopted from [154]

The view of the RA was used to generate the corresponding view of the AA, as seen in Figure 4.2. Figure 4.3 depicts the procedure followed for this derivation. For each view of the reference architecture, this approach was used; the application's necessary entities were first listed; then entities from the corresponding RA view were chosen based on the entity from the application. The required entity is reused if it could be identified in the RA; otherwise, a new entity was introduced to the AA. If the entity is located in the RA, it was examined to see if it can be reused in its original state or whether it has to be changed. If the names of the modules were the same or if the modules were interchangeable (e.g., financial management vs. economic management [155]), the modules were considered entirely reusable. The module may be composed or decomposed if it is not reusable in its present state and thus had to be modified. In a composition, multiple RA modules were merged into a single AA module. As an example of a composition, a data transfer module and data collection module could be merged into a data processing module (see [156]). After the decomposition, an RA module is broken down into several smaller modules in the AA. Finally, the reusability of the RA's entities was explored and the RA's usability (to derive the AA) considered. Concluding, making an AA for a particular settings (i.e. case study), serves as validation for the RA [139, 154].

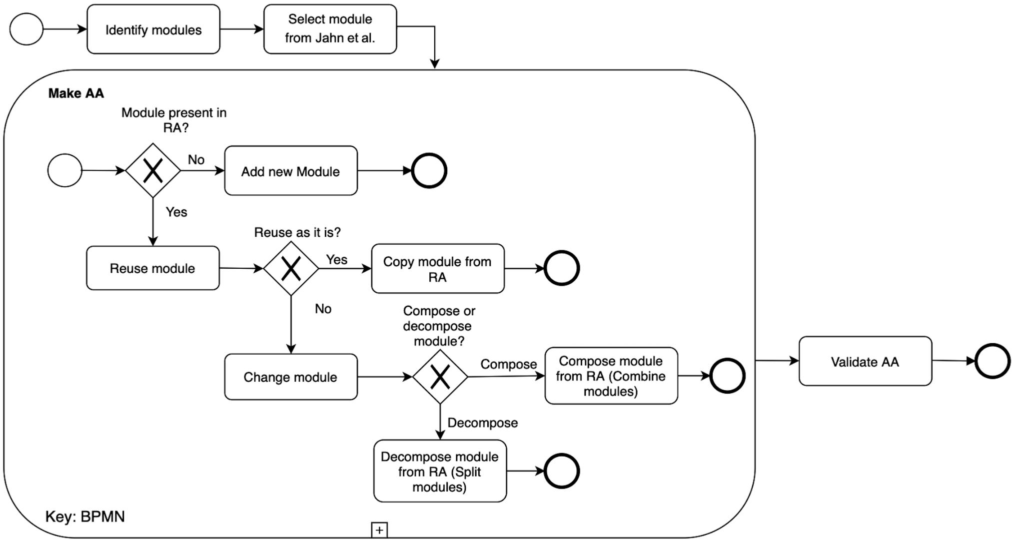


Figure 4.3: Approach followed in building AA from RA Adapted from Tummers et al. [154]

This approach as described above and depicted in Figure 4.3 , has been used in a variety of other domains such as agriculture [154]supply chains [157], and smart warehouses [151].

4.3 Results

4.3.1 Domain analysis

To scope and model the domain, we performed a systematic literature review [158] to identify papers in which HISs, their domains, stakeholders and, concerns and features were described. This resulted in a set of 11 papers [59, 72, 133, 159–166].

Domain scoping

HISs cover a wide range of sub-domains within healthcare. Many HIS papers focus on the hospital sub-domain [59, 160, 165], others focus on the primary care [133], pediatrics [161], outpatient care [163, 165], and diabetes care [166]. The most common stakeholders and their concerns for HISs development are presented in Table 4.1. While some stakeholders are generic for HISs, such as the patient, other stakeholders are more domain-specific, such as the Laboratory.

Table 4.1: Key stakeholders (in alphabetical order) and their main concerns.

Role	**Concerns**
Administrative staff	Wants easy data entering and retrieval
Automated data source	A protocol to safely upload data from heart rate monitor, wearable technology, medical robots, et cetera
Care professional	Wants system to be easy to use such that information can be quickly entered, retrieved, and shared
Government	Wants the system to comply with all their regulatory standards.
Healthcare manager	Needs system to provide overviews and reports.
HIS developer	Develops system in time within the planned budget
Insurance company	Wants compatibility with their system for reimbursement.
Laboratory	Wants compatibility with their measurement devices
Other HIS	Needs to be able to communicate with HIS and exchange data
Patient and/or representative	Wants data to be stored safe and secure. Wants care professionals to have the right information at the right time. Wants reimbursement of care.
Pharmacist	Needs medication management to be an integral part of the system
Plug-in developer	Wants easy to use platform for plug-in development
Research institute	Needs system to provide structured data such that it can be used for research.
Secretary	Needs system for making appointments and administrative tasks.
HIS administrator	Wants system that is easy to maintain and adequately documented.

Domain modeling

To model the features of the HIS domain, feature modeling was adopted. A feature model represents the domain knowledge and desired system by distinguishing common, alternative, and optional(e.g., sub-domain specific) features of the system, and the interdependencies amongst these features [167]. The feature is defined as ″a prominent or distinctive user-visible aspect, quality, or characteristic of a software system or system″ [168]. Sub-features of a more general feature are shown under the most general feature in a tree-shaped model [168]. Our feature model for the HIS is presented in Figure 4.4 and is based on the features mentioned in the literature.

We split the full set of features into six main features (Middle column of Figure 4.4). The Generic Management Information System (MIS) feature contains non-domain-specific features. The Data management feature contains features related to the management of data and data-driven decision-making by care professionals. Medication management and Patient monitoring are typical HIS features. Planning & scheduling is a feature mainly used by secretaries and administrative staff. Last, but certainly not least, the Security feature must ensure the system's resilience and protection of its data.

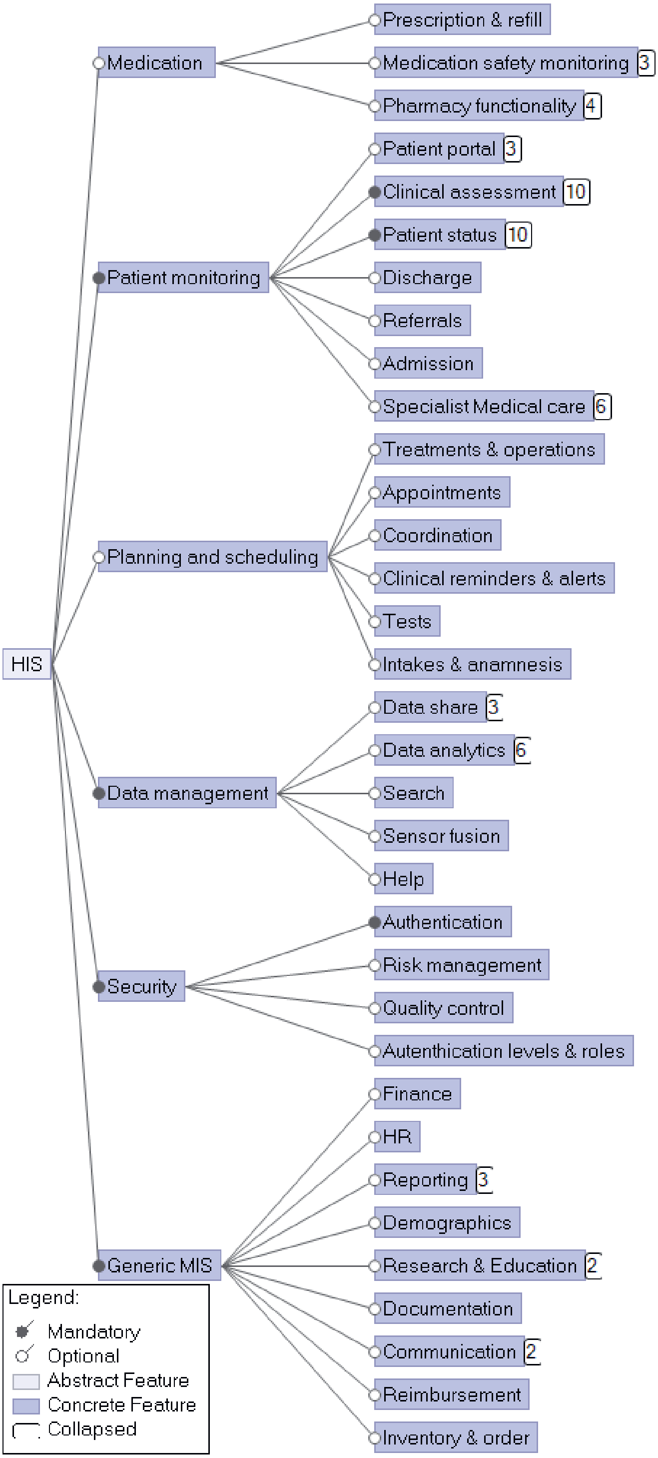


Figure 4.4: A downsized version of the Feature model for HISs. Numbers on the right-hand side of the features represent the number of sub-features not shown.

4.3.2 Architecture design

In the next section, the selected viewpoints were used for designing the RA are described. In the subsequent sections, the HIS RA views were built from these four viewpoints.

Selection of views

Although the HIS's main purpose is to assist in the current daily operations, it should also be flexible and adaptable to facilitate different long-term visions [136, 169]. To do so, the RA needs to cover all features of the feature diagram in Figure 4.4. The RA should also cater to users in all different sub-domains of healthcare and facilitate tailoring to local needs. After all, a hospital HIS needs to meet different demands than a general practitioner's HIS and thus, will have different architectural decompositions.

For modeling the RA, we adopted the Views & Beyond (V&B) approach [153]. This approach consists of selecting out of 17 predefined viewpoints the ones of interest to certain stakeholders. The four viewpoints of particular interest to key stakeholders in the healthcare domain selected for modeling the HIS RA are the context diagram, decomposition view, layered view, and deployment view.

Context diagram

The context view of a system contains the entities that are outside the system's scope but have a direct relation with the system [170]. The context diagram represents the context view and shows the system boundaries, environment, and the entities it communicates with [171]. The reference context diagram for the HIS is presented in Figure 4.5.

The external entities and their communications with the HIS were based on the stakeholders and their concerns from Table 4.1. Six external entities are considered obligatory: the HIS cannot function without them. The optional entities can be absent in simpler HISs such as automated data sources or are (sub)domain-specific, such as the laboratory. Besides, some (sub)domains may require specific entities that are not shown in the reference context diagram. Many entities have two-way communication with the HIS, meaning that the HIS communicates with the entity and vice versa. External entities with a one-way communication with the HIS, are rarer. For example, a governmental organization can receive reports from the HIS, but this organization has no authorization to access the HIS data. We only describe one type of communication per interaction due to space limitations, in practice, there are many more possibilities.

Decomposition view

The decomposition shows how a system can be decomposed into multiple (sub)modules and how they relate to one another (parent-child). This view often is the basis for HIS design, development, and system documentation [172]. The decomposition view helps to check for the presence of the required modules for all stakeholders. The HIS RA decomposition view consists of six modules with 34 sub-modules, see Figure 4.6.

The first module is Medication management, containing sub-modules related to medication handling, distribution, and safety monitoring. The second module is Patient monitoring and contains sub-modules related to the assessment, admission, discharge, status, and referrals of patients, and is input for the electronic health record. The Patient monitoring module also contains a sub-module labeled Patient portal in which the patient can check his/her files. The Security module with the sub-modules Authentication, Authorization, and Security mechanisms must ensure the privacy and security of the HIS and its data. Module number four is Planning and scheduling, with sub-modules used by various stakeholders to ensure proper care coordination. The Generic MIS module is not healthcare specific. Its sub-modules are important to keep track of assets, such as staff and inventory, to provide means for organization-wide communication, quality control, and financial affairs. Often the features from the generic MIS module can be found in so-called enterprise resource planning (ERP) systems. These systems are business management system solutions which are used for managing, automating, and integrating all the business functions within an organization [173–175]. These five modules generate data, which needs management. This happens in the Data management module, which has sub-modules to ensure proper import, sharing, analysis, and data search.

Figure 4.6 shows all described modules and sub-modules of the RA for HIS. A specific Application Architecture (AA) consists of a selection of these modules tailored to the stakeholders' requirements.

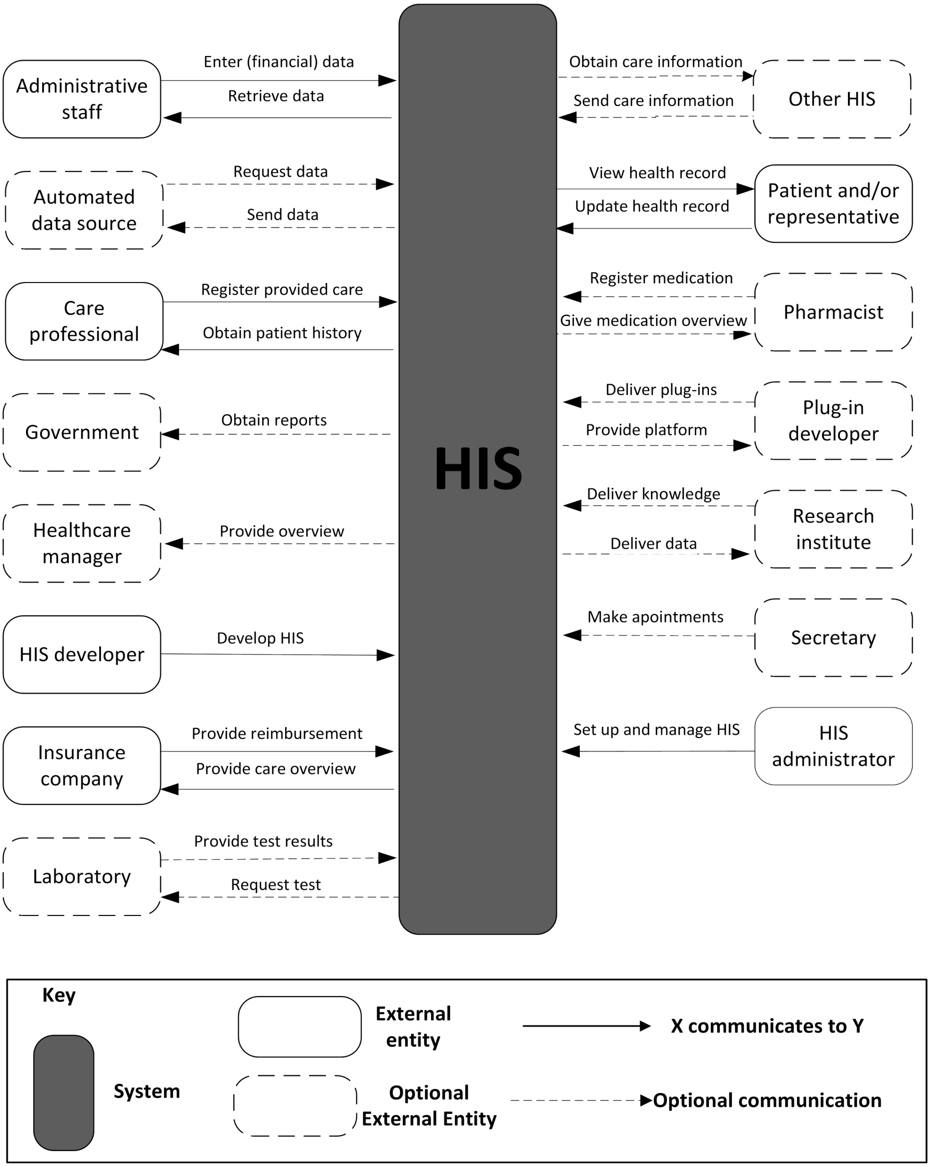


Figure 4.5: The reference Context Diagram. Only the interactions considered the most important are shown.

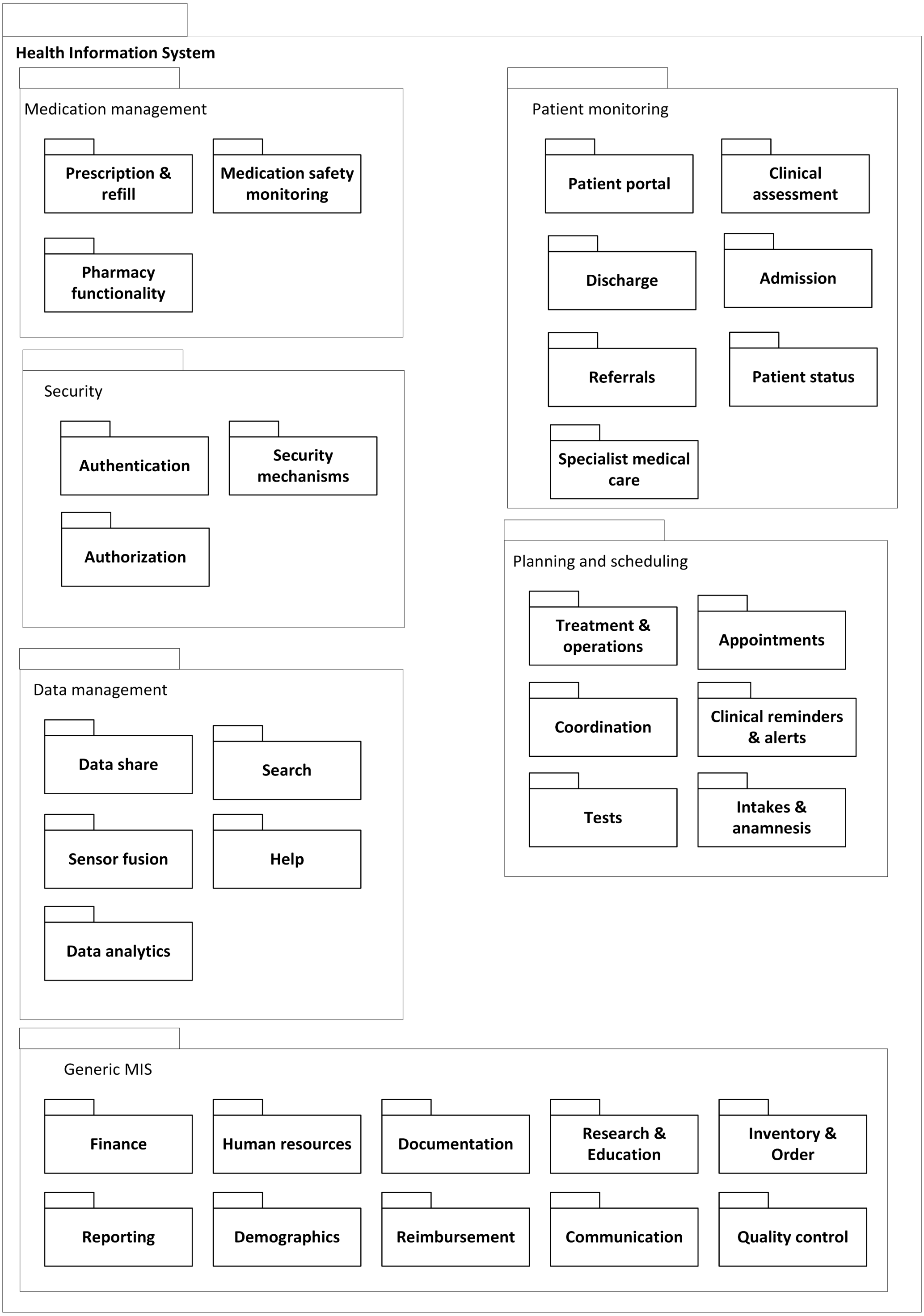


Figure 4.6: The reference decomposition view of the HIS.

Layered view

The layered view reflects the software modules' allocation into different layers, based on a unidirectional "allowed to use" relationship between the layers Clements et al. [153]. We decided to base our layered view on the standard of enterprise software systems because of its flexibility (Figure 4.7). Starting at the top, the layered view consists of a presentation layer with a User Interface (UI). The presentation layer relies on the business logic layer that determines how data are created, stored, and processed. The business logic layer contains the Planning and scheduling, Generic MIS, Patient monitoring, and Medication management modules from the decomposition view (Figure 4.6). These four modules, the backbone of any HIS, generate and use data from the Data management layer. The Data management layer contains sub-modules to simplify access to the data. To provide overall HIS security, a vertical layer connected to all three horizontal layers was added. This Security layer contains the modules: Authentication, Security mechanisms, and Authorization for safety and security at all system layers.

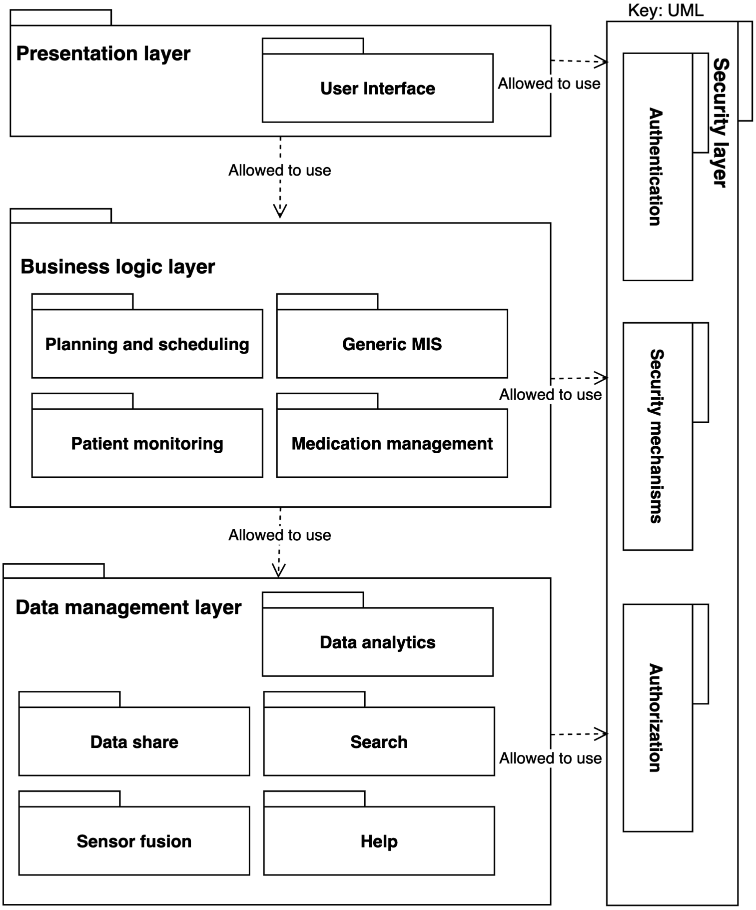


Figure 4.7: The reference layered view for the HIS.

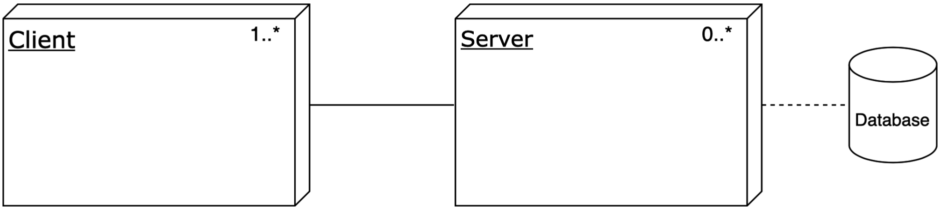


Figure 4.8: The reference deployment view of the HIS.

Deployment view

In the deployment view, software modules are allocated to the hardware entities on which they are executed. This view is useful for analyzing the performance, availability, reliability, and security aspects of the system [153]. Due to the vast diversity of HISs, we decided to develop a generic deployment view that can represent many of HISs across care domains

The deployment view (Figure 4.8) shows one or more clients and zero or more servers. If there is a client only, and no server, the deployment is a standalone desktop application or a thick-client with all modules on the client-side. A client-server application consists of at least one server and multiple clients, for example, thin clients with modules located on one or multiple servers. Finally, a system with multiple clients and multiple servers that communicate using cloud computing technology is cloud-based.

A system will most likely have a back-up server in case the original server goes down, but a combination of other types of servers is also possible. These other types of servers could include load balancing servers to allow for big data analytics, as well as application servers, web servers, and database servers. The RA deployment view also provides space for a web-based application. In that case, only an internet browser is required on the client-side with which the end-user can use the HIS. The server is often provided by the software supplier, which contains the modules to host the web page and store the data.

Depending on the specific requirements the allocation of the modules as identified in the decomposition view can be allocated in various different ways over the selected nodes in the deployment view.

4.3.3 Case study Chiba University Hospital

This study's primary objective was to propose and evaluate the RA. We decided to base the illustration and evaluation on a case study from the literature, as no site visits were possible due to the COVID-19 pandemic. For our case study, we used the well-detailed article by Jahn et al. [176] in which they compare a Japanese and German hospital HIS using the three-layer graph-based meta-model (3LGM^2) [149]. When presented and inspected visually, the 3LGM^2 model combines the UML decomposition view, uses view, and layered view.

Our case study was done by developing an AA. for the Japanese Chiba University Hospital (CUH) based on Jahn et al. [176]. Figure 4.3 shows the approach followed to build the AA.

Feature diagram: The 104 modules of the CUH model in Jahn et al. [176] (page 6 Figure 5) were mapped onto the features from our feature module (Figure 4.4). We added 47 sub-features to meet the level of detail presented in Jahn et al. [176]. Interestingly, Jahn et al. [176] listed more detailed Patient monitoring and Generic MISs features, which we included as sub-features in Appendix Figure 1. In contrast, our RA was more detailed concerning the other HIS features.
Context diagram: Although the stakeholders of the HISs are not explicitly mentioned in Jahn et al. [176], we were able to make the application context diagram based on mentioned systems such as a Laboratory Information Systems and a Pharmacy Department System. The stakeholders and other entities of such systems combined with the obligatory entities and interactions from Figure 4.5, provided the application context diagram. There was no need to add extra external entities (see Figure 2 in the Appendix). We used 12 out of 15 (80%) entities from the RA context diagram.
Decomposition view: The decomposition view extracted from Jahn et al. [176] is presented in Appendix Figure 3. Despite slightly different wording in the labels of (sub)modules, we could make the decomposition view, which listed 104 modules from the feature model. Seven sub-modules from our RA were not found in the Japanese HIS and removed from the application decomposition view. Therefore, we utilized 27 out of 34 modules from the RA decomposition view, resulting in a re-use of 79%.
Layered View: Although the authors used the term " layers" differently than we do, the provided information allowed us to derive the layered view using our own design choices. The result was the same as depicted in Figure 4.7 above.
Deployment view: Jahn et al. [176] provide limited information about the CUH HIS deployment, but it does show the CUH databases. From this information, we inferred the deployment situation at the hospital. The deployment view is available in Figure 4.10.

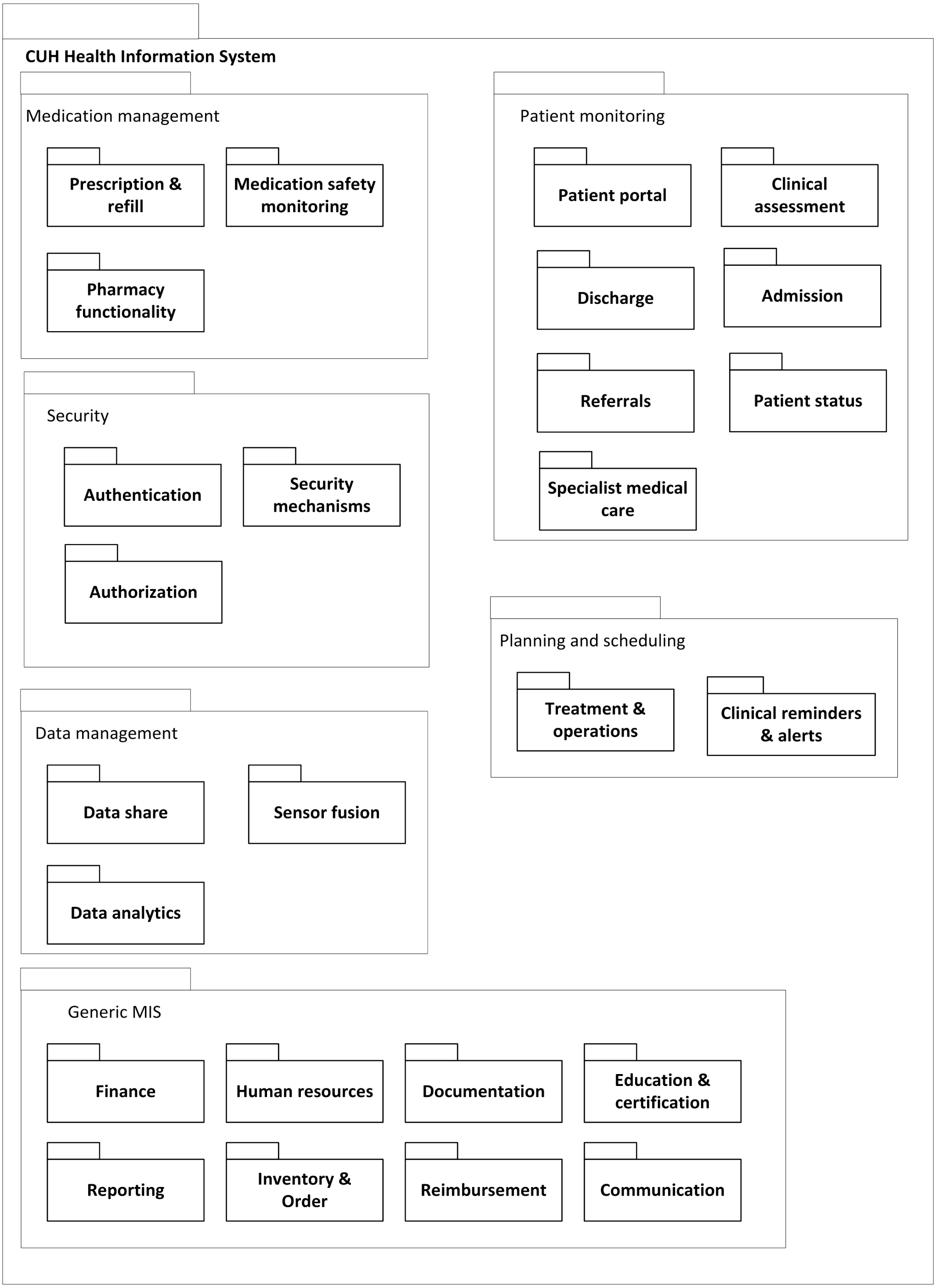


Figure 4.9: Decomposition view for the Chiba University Hospital based on Jahn et al. [176].

As discussed above and shown in the Appendix, we could successfully derive an AA for the CUH case from our RA. Making the views for the case study took us about two days (16 hours). Based on this case study, we made some minor changes to our RA, which were already included in Figures 4.4, 4.6, and 4.7.

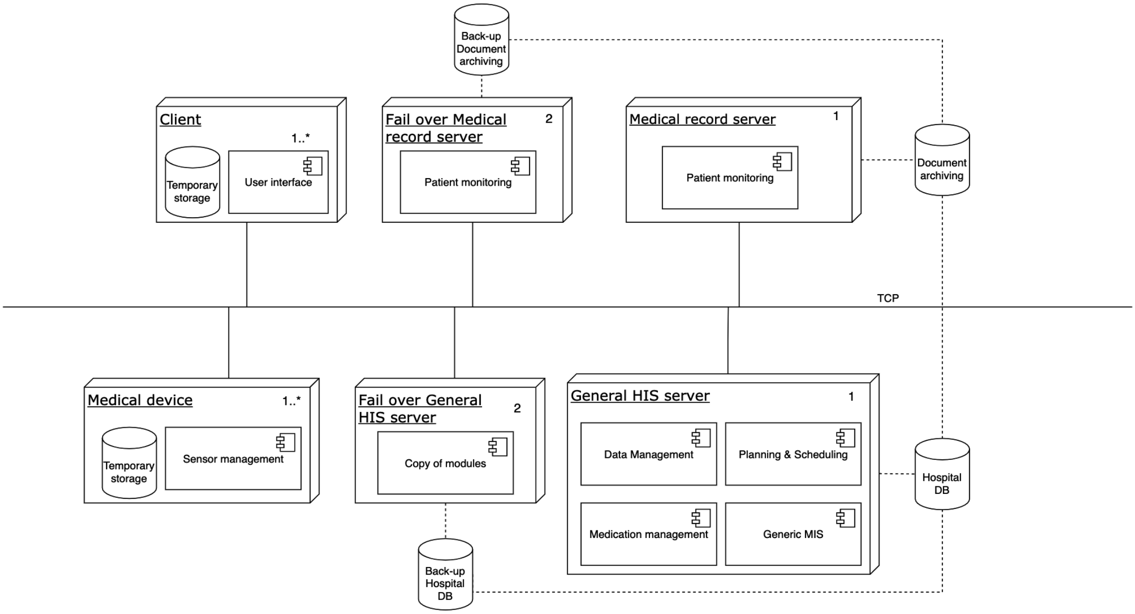


Figure 4.10: Deployment view for the Chiba University Hospital based on Jahn et al. [176].

4.4 Discussion

To the best of our knowledge, this is the first RA for the health care domain built using standard architecture design approaches from the software architecture community. In this discussion, we critically reflect on our results and compare this study with related work.

4.4.1 Critical reflection on the results

For the domain analysis, we relied on scientific articles. A more extended data collection from grey literature or expert interviews might have yielded different input for the viewpoint selection. We believe that the scientific articles provided a factual basis for the viewpoints because of their diversity across care domains, and, indeed, our case study did not suggest otherwise.

Based on the domain analysis, we identified 15 key stakeholders for HISs because of their relevance to almost all HISs. The domain was modeled with the feature diagram, which provided a broad overview of the different features demanded for HISs. The feature diagram included the most relevant features and, when needed could be extended with additional (sub)features, as illustrated in the case study Section. This allows the feature model to evolve with the changing health care domain.

Based on the key stakeholders' concerns and input from the domain analysis, four viewpoints were selected to model the RA. Together, these viewpoints gave a broad and solid overview of HISs. The Context Diagram and Decomposition View showed the architecture from the stakeholders' perspective, the Layered and Deployment View provided a standardized technical representation of HISs. The Deployment View (Figure 4.8) was modeled generically to allow for various deployment alternatives, as illustrated in the case study (Figure 4.10).

In current practice, the modules described in the decomposition view are often implemented by a combination of systems. In a hospital, for example, a hospital information system, an order management system, a pharmacy information systems, and many more systems are used. At first sight, a fully integrated ERP system would be an option to align these processes and systems. However, there are several difficulties in using ERP systems in the healthcare sector. First of all, the alignment of business processes with the ERP system is not an easy task, and the success of the project, therefore, depends on the complexity of the processes in the environment. For this alignment, either the processes or the ERP system have to be adapted, but some ERP systems require a lot of effort to be adapted to the required processes. Another problem is related to the vendor lock-in problem [177]. When an ERP system is adapted for the healthcare provider, there is too much dependency on the vendor and the consultants who can provide the required services.

The case study was based on a peer-reviewed article due to the COVID-19 pandemic. Although Jahn et al. [176] did not explicitly name stakeholders, the paper contained sufficient detail to derive the context diagram and the decomposition view. Similarly, we were able to derive the layered and deployment view based on the detailed information Jahn et al. [176] provided. The use of four views to derive the AA was demonstrated. In theory, the same procedure can be used to generate other potential perspectives (e.g. use views, layered views etc.). To do so, the appropriate views must be defined based on the chosen system's particular application requirements [153] . Till all the necessary views have been determined, the approach described above will be followed; that is, reference views will be established first, followed by application views. When using this reference architecture for the development of a new system, it is very important to make use of the different standards for HISs. In order to ensure interoperability with other systems, standards such as HL7 FHIR should be used[109] Furthermore, the diagnoses in the systems should also be standardized using codes such as ICPC-2 or ICD11 [110, 111].

Future work will expand towards cases in the long-term care domain to further demonstrate our RA's applicability.

4.4.2 Related Work

Several other RAs for HISs have been published. The pioneer RICHE RA from 1993 [141] has an open architecture with three layers: user applications, basic applications, and information systems. Despite its old age, the paper described many problems that have remained unsolved up until today. Wartena et al. [142] described in 2010 a RA for a personal telehealth ecosystem with a focus on networking and communication, ignoring other features.

More RAs for HISs are found in grey literature, such as white papers and technical reports. These RAs are often characterized by none [143] or some diagrams only [144–146], and do not apply any formal software architecture modeling technique, as defined in the computer science literature [178].

We found three papers that used diagrams systematically to describe their RA for hospitals [147, 149], and healthcare in general [148]. Nictiz [147] presented an RA for hospitals using an Archimate model [179]. Their RA showed similarities with ours: their domain '*reference model*' contained many elements from our decomposition view and layered view. However, the Archimate Model is limited to the scope of enterprise modeling [180] and is based on the by now replaced IEEE 1471 standard [181]. In contrast, UML has a much broader scope and contains many more modeling concepts to choose from, 150 instead of 50. Winter et al. [149] based their RA for the hospital domain on the UML-based 3LGM^2 model, which had also been used by Jahn et al. [176]. Winter and colleagues' metamodel for modeling Hospital Information systems, shows similarities with our RA as explained in Section 5. An RA with a similar scope to ours is ATOS' "IT Reference Architecture for Healthcare" [148]. They did not use UML models, but an informal approach to display the ICT services for HIS development. Their RA shows some overlap with our decomposition view but ignores a deployment view.

Compared to the other RAs, our RA is generic, uses UML models, and addresses the entire healthcare domain.

4.5 Conclusions

In this study, we showed that the methods of the software architecture design community could be used in the healthcare domain effectively: we proposed a generic RA for HISs. We have shown the suitability of the RA for deriving the AA for a University hospital in Japan. Our method of evaluating an RA was successful for one case study. In our future work, we will use this method to a greater extent and apply the reference architecture for designing the architecture of various other HISs.

Appendix

Application architecture Figures

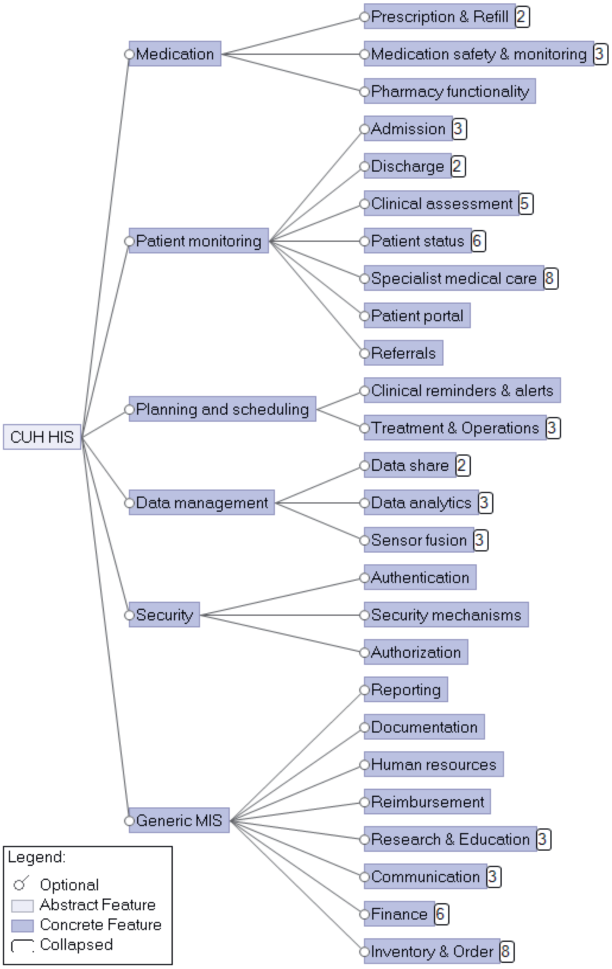


Figure 4.11: Downsized version of the feature model for the Chiba University Hospital based on Jahn et al. [59]. Numbers on the right hand side of the features represent the number of sub-features shown.

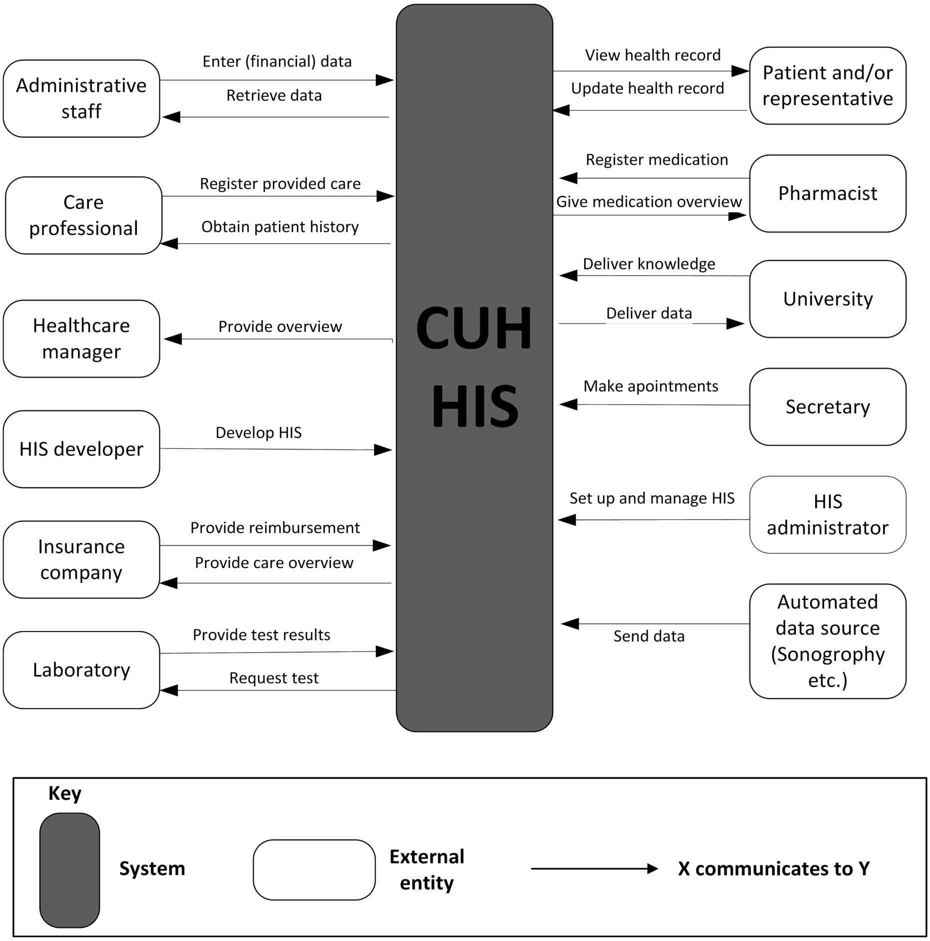


Figure 4.12: Context diagram for the Chiba University Hospital based on Jahn et al. [59]

Study 4: Big data analytics system design

5

Introduction: Care providers in intellectual disability care use a wide variety of health information systems to register their provided care. This leads to a large set of structured and unstructured data with a high potential for research that cannot be utilized in the current state. Therefore, a big data system needs to be created to fully utilize the potential of this data collected in intellectual disability care. An important challenge in this context is the design of a proper architecture for big data analytics dedicated to intellectual disability care. We, therefore, need to know what the health information system architectures in intellectual disability care look like and what their commonalities and variabilities (differences) are. Furthermore, the consequences for the big data analytics architecture of these health information system and privacy and security issues need to be identified.
Methods: We used a multiple-case study approach to identify the architectures of four health information systems in the Netherlands. We interviewed seven stakeholders from four unique HISs and gathered multiple documents regarding the system infrastructure. Similarities and differences between the systems and the main challenges regarding the privacy and security of a big data analytics system were identified.
Results: Four architectural views on the health information systems were derived: a context diagram, decomposition view, layered view, and deployment view. We discuss multiple important aspects of security and privacy of a big data analytics system for intellectual disability care and present a general architecture design.
Conclusion: Multiple challenges must be solved before the full potential of data in intellectual disability care can be unlocked. This research provides first guidelines on how these problems can be overcome and provides guidance towards developing a big data analytics system.

This chapter is based on an article submitted to an international journal: Big data for intellectual disability care: architecting Dutch health information systems. J. Tummers, H.Tobi, C.Catal, B. Tekinerdogan, B.Schalk, G.Leusink

5.1 Introduction

In recent decades we have seen massive digitization of healthcare [79]. This is even more so when both medical and daily care data are registered, such as in the care of people with Intellectual Disabilities (IDs). Different types of data are stored in a complex landscape of Health Information Systems (HISs), as different care providers use different HISs. In addition, due to this care being long-term, HISs may also change over time [182].

The data captured in this landscape of HISs is nowadays mainly used for client and patient monitoring. In addition, this data has great potential for researching health over time and for natural history studies amongst this vulnerable population [20, 21]. This potential can help fill the gap caused by the usual underrepresentation of people with ID in medical studies [24]. Traditional data management tools and technologies could not keep up with the massive amounts of data being generated and the requirement to acquire, store, analyze, and process it in a reasonable amount of time [183]. Therefore, technology for big data has been developed, such as cloud computing allowing large scale analytics and artificial intelligence [183, 184]. Big data is more than a lot of data; big data is defined by the four Vs: Volume (amount of data), Velocity (speed of data), Variety (types of data and sources), and Veracity (reliability of data). So, to fully utilize data collected in ID care, a big data system needs to be created.

There are multiple challenges to overcome before this routinely collected data from various HISs can be stored and re-used for big data analytics. One key technical challenge is the design of a proper architecture for a big data analytics system that allows the inclusion of data from multiple HISs. An architecture is the description of a system or collection of systems in several diagrams with text (so-called 'views' [185]). Such architecture needs to allow for the integration of the different data sources into one big data analytics system in the presence of different interoperability issues [186]. Other challenges are related to stakeholder concerns on privacy and security [182, 187].

A study amongst 328 care professionals involved in ID care reported 52 distinct HISs in The Netherlands [182]. Even within one single ID care organization, at the very least two HISs are used: one with electronic client records used by daily care providers and one with electronic patient records used by physicians. Moreover the electronic client dossier system may be related to an external sheltered workshop record and the electronic patient dossier system to the electronic pharmacy system, the hospital system, and that of, for example, the physiotherapist and the psychiatrist. All these HISs are different, which is at least partly the result of the lack of widely accepted, let alone standard use of HIS architecture [186], further complicated by the diversity of health sectors involved in ID care.

Consequently, a system to analyse the data on people with ID not only needs to ensure privacy and security but also needs to accommodate the software architectures of this range of HISs. Essential input for the building of a big data analysis system are the commonalities and variabilities (i.e. differences) in the architectures of these distinct HISs [188]. Therefore, the aim of the present study is to map the to-be-built architecture for a big data analytics system based on a multiple-case study on the architectures of HISs in different stages of development and acceptance used in ID care. Such that better re-use of the routinely collected data in ID care can be made. To reach this aim, we have formulated the following research questions:

- RQ1 What do the HIS architectures look like and what are their commonalities and variabilities?
- RQ2: What consequences do the HIS architectures and the privacy and security issues have for the architecture of a big data analytics system in ID care?

The paper is organized as follows: in the next section, the research methods used for the multiple case study approach and the derivation of the big data analytics system architecture are presented; in Section 5.3, results are presented; Section 5.4 shows the discussion and, finally, Section 5.5 contains the conclusion.

5.2 Methods

The study consisted of two parts: a multiple-case study on the HIS architectures which feeds into the second part, the design of the big data architecture also addressing privacy and security issues for data re-use.

5.2.1 Case study research

To design a big data analytics system for ID care, it is crucial to model the architectures of HISs. We based our case studies on the guidelines on case study research in software engineering from Runeson and Höst [189]. We decided to include four cases on HISs that cover a wide variety of the HISs used in ID care (Table 5.1). Throughout the report, we identified these HISs as case A, B, C, and D. We based this order on the age of the system (old to new), and the setting (short-term-care in general practice, to long-term care for people with an ID). Please note that Case D is a prospective case study with a system that is currently under development. We selected the cases through our academic collaborative *"stronger on our own feet"* * (A & C), contact with professional associations (D), and by contacting a healthcare organization that was still unknown to us (B).

For each of the cases, we had multiple sources of information (Table 5.1) including, for the existing cases (A,B&C), semi-structured interviews with someone from the software provider and from the care organization. The list of (translated) questions is presented in the Appendix. Each interview lasted approximately 90 minutes. After the interview, the outcomes were put in a short report. This report was discussed with the interviewee(s) about two weeks after the interview and adjusted where necessary. Due to the Covid-19 (SARS-CoV-2) pandemic, six interviews took place online. No interviewee objected to the online session being recorded. One face-to-face interview was audio-recorded with permission. The interviews were carried out between March 2021 and January 2022.

Table 5.1: Information sources per case

	Case A	***Case B***	***Case C***	***Case D***
System Type	EPD	EPD	ECD	EPD
Case study type	Retrospective	Retrospective	Retrospective	Prospective
Interviewee 1	ICT developer	System administrator	System administrator	Two ID physicians
Interviewee 2	System initiator and developer	Product owner / mid-level tech manager	Software Support Specialist	-
Manual received	Yes	No	Yes	No
Other information	Overview with diagrams	HIS interactions Screenshots	Database overview	Brief of requirements

---

* A collaboration between the Radboud university medical centre and sevenlager care provider organizations for people with intellectual disability `https://www.sterkeropeigenbenen.nl/english`

Case A is the general practitioner information system TransHis [190], which is also used by ID physicians. The ID physician is trained to provide care for people with intellectual disabilities and has central role in the Dutch organization of ID care, which is divided in people living at home receiving mainly care from a GP, and in people living in institutionalized long term ID care. We interviewed an ICT developer from Radboudumc who also works with the data from TransHis and one of the initiators of TransHis. Case B is the electronic patient dossier Ysis, developed by Gerimedica. We interviewed a mid-level manager / product owner from Gerimedica and a system administrator working at a large care organization for people with an ID in the west of the Netherlands. We also received documentation about the interactions of Ysis with other HISs and some screenshots from the system administrator. In institutionalized Dutch ID care, Case C is a frequently utilized electronic client dossier. The interviewed system administrator worked at a care organization for people with ID with multiple locations in the east and south of the Netherlands. Furthermore, we received access to the online manual from the software supplier. Documents received were the manual and a document containing the history and the database description. Case D is an HIS under development by the Dutch association of ID physicians (NVAVG) named AVG-IS. We interviewed two ID physicians initiating the system and responsible for the development of this HIS. We also received their brief of requirements for the AVG-IS. For each case all sources of information were considered when making the diagrams and describing the systems in the results section.

Architecture derivation

We based the architecture derivation on a reference architecture [186] that consisted of diagrams from the Views & Beyond approach [185]. We selected the four views that allowed us to describe the HISs from the contextual level with the system's interactions with external entities, to the more technical description with the deployment of software modules on servers. For further information on the views used, see Table 5.2.

Table 5.2: Summary of views used

	Context diagram	*Decomposition view*	*Layered view*	*Deployment view*
Type	Overview	Module style	Module style	Allocation style
Shows	Entities outside the system's scope but directly related to the system [170]	Shows how a system can be decomposed into multiple (sub)modules and how they relate to one another (parent-child). [172]	Reflects the software modules' allocation into different layers based on a unidirectional "allowed to use" relationship between the layers [185]	How software modules are allocated to the hardware entities on which they are executed.
What's it for	Depiction of system whose architecture is documented [171]	Basis for HIS design, development, and system documentation [172]	Promote modifiability and portability. Facilitate communication of code-structure to developers. [185]	Analyze the system's performance, availability, reliability, and security [185]

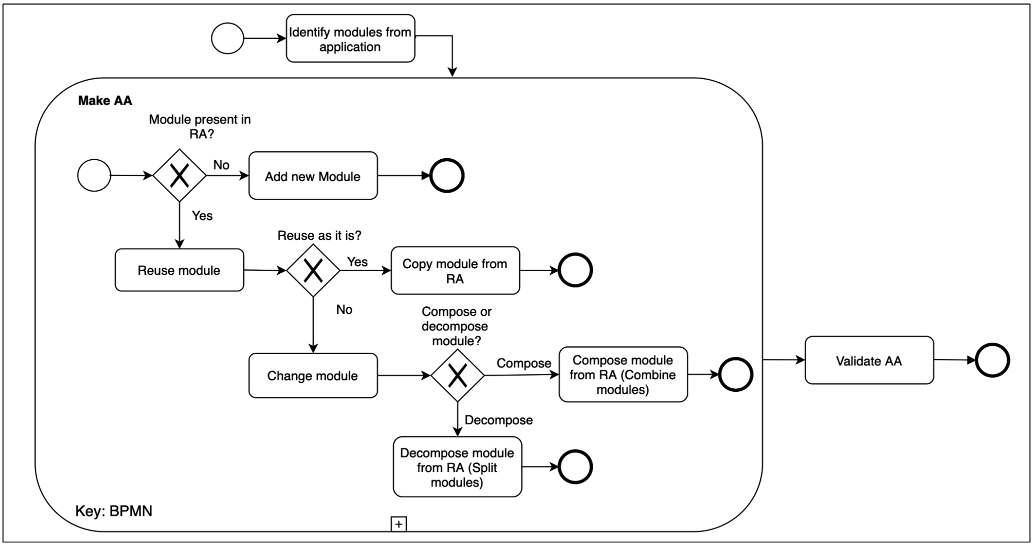


Figure 5.1: Approach to building an information system's architecture from the Reference Architecture Adapted from Tummers et al. Tummers et al. [186]

To derive these views, we used the approach shown in Figure 5.1. We looked at whether we could map the obtained diagrams and views from Table 5.2 for each of the cases on the RA [186]. Large inconsistencies between information sources would have been discussed with the interviewees until a consensus was reached. However, large inconsistencies did not occur. Minor inconsistencies will be addressed in the results section. If an interviewee mentioned some entities or modules that were not one-on-one available in the reference context diagram, the authors (1 and 2) discussed these and tried to fit them in the RA. If the entity or module could not be fitted, we added an extra entity or module to the diagram and indicate this as such in the figure.

5.2.2 Big data architecture

Based on the outcomes of the cases, we built an architecture of a big data analytics system for the care of people with intellectual disabilities. This system should be able to handle the data from all the different HISs in ID care. The architecture for the big data analytics system is based on the guidelines presented in [191]. Furthermore, we had a closer look at the privacy and security issues involved and presented solutions for them drawn from existing literature.

5.3 Results

We start with a brief overview of the results obtained from all information sources, particularly the interviews (Table 5.1). Then, we synthesized the four case studies into four diagrams: the context diagram, decomposition view, layered view, and the deployment view. Based on these elements, we derived the big data architecture.

5.3.1 Information sources from cases

Case A

HIS A is mainly used as a GP information system in the southeast of the Netherlands. According to interviewee 1 (Table 5.1), TransHis makes sure to structure and code the information as much as possible, such that it can be re-used for research. Examples of this research include [192], [193], and [194]. According to the system initiator, the re-use for longitudinal research was one of the main reasons for the HIS development; it facilitates exporting structured fields anonymously. According to the ICT developer, the necessity to register the "reason for encounter" (visit, consultation) is crucial for facilitating research with the systems' data. After all, the reason for encounter may be something else than a symptom. Furthermore, everything in the HIS is linked to a diagnosis (e.g., diabetes, laryngitis).

The overview with diagrams of the system and the manual showed that the system is straight-forward, and functional, with a simplistic-looking user interface. The interface requires a minimum amount of clicks for the daily end user. From the researcher's perspective, the database is well-structured and facilitates the FAIR principles [195].

Case B

According to the product owner, HIS B, Ysis, is an electronic patient dossier in chronic care and is rarely used as a stand-alone system. According to interviewee 1, Ysis was used by ID physicians and out-of-hours general practitioners (GPs), whereas there are in total seven different HIS in the care organzation. From the document with the HISs' interactions at the care organization, we noticed that patient/client level data is exchanged automatically between Ysis, the electronic client dossier and the electronic prescription system only. The product owner stated that Ysis could be linked with most major electronic client dossiers via Application Programming Interfaces (APIs). Furthermore, there is a Single Sign-On (SSO) link with a major electronic prescription system. Ysis is organized per patient. Each care professional registers according to her/his professional guidelines and has access to what she/he registered for the own patients only. Professionals can only access data from other professionals and other patients if provided with the appropriate authorization. The anonymized data from Ysis is used for research mainly by one university and for policy-making by multiple institutions [196]. According to the product owner, the wide variety of parties interested in the data could influence the patients' willingness to participate in research.

Case C

Case C is a generic electronic client dossier suitable for multiple types of long-term care, including ID care. The system was originally developed for administrative purposes (e.g., hours worked by professionals) and has been expanded to include daily reports by personal care aides on clients and a client portal. The client can view see her/his own data, which includes care dossier data, reports and measurements about the personal care pathway. The system administrator explained that multiple roles are available to care professionals in the HIS, each with its authorization level. The HIS is accessible as a web application and works with two-factor authentication and Single Sign-On to (external) systems. The software support specialist and system administrator mentioned a standardized API available to allow for integration with other systems. Via this API, there may be, for example, a connection between an incident registration system and a staffing information system. The document with the database overview (>500 pages) showed that a wide variety of data entry fields is stored in the database, including bank account numbers and which professional cares for which client. The system administrator stated that this HIS is one of the five applications used in the care organization. It has been acknowledged that this wide range of systems adds complexity and confusion to the care providers. In this care organization, data export was only used to fulfill legal requirements (e.g., COVID vaccination registration). Interviewee 2 confirmed that the possibility of exporting data via Excel is not used for research.

Case D

HIS D is to be designed for ID care specifically. This HIS is developed because currently used GP information systems cannot address the environment of ID care characterized by long-term care involving a range of both medical and non-medical professionals. Therefore, the interviewees aim not only at ID physicians for end users but other professionals who deliver medical care for people with an ID, such as paramedics and psychiatrists. Daily care providers ought to be able to access an automatically generated summary in HIS D from their electronic client system. The interviewees would like the database behind HIS D to facilitate research. The brief of requirements contained the business processes behind the workflow of the end user and the entity-relationship (ER) diagrams. From the brief, we concluded with which other existing HISs and other systems to be designed should be able to communicate.

5.3.2 HIS architectures

Context diagram

We describe the context diagram for the four case studies in Figure 5.2. Nine entities appeared in all four cases. Six entities were present one to three times. One entity was added compared to the reference context diagram: a messenger service. The interviewees mentioned that some entities were not one-to-one available in the reference context diagram but could be mapped on one of the existing entities. For example, a logistics system (case B), a system for communication with insurance companies (Case D), and a link with a professional association for medical guidelines (Case D) were considered as external plug-ins for Figure 5.2.

There were some small inconsistencies between the interviewees; we decided to show the entity in Figure 5.2 if at least one interviewee mentioned it. Five times the interviewees working at the software supplier mentioned an entity that was not purchased and/or used in the care organization. The care organization's system administrator from Case C mentioned an external governmental entity, while this entity did not appear during the interview with the software support specialist. Case B's system administrator identified the laboratory as an external entity, whilst the product owner did not identify this.

Decomposition view

The decomposition view for the four case studies is presented in Figure 5.3. Each case consisted of five main modules. Eighteen sub-modules were used in all four cases and seventeen sub-modules were available in one to three cases. Two sub-modules, which were present in [186] (Human resources and inventory & order) were not available in any case. We added another module, which was considered of great importance in case D: coercive care act. This is designed to help meet the Dutch legal to register coercive care and evaluate the measures taken. Some modules came from the interviews, which were not one-to-one available in the reference. For example, a care plan was mentioned as a sub-module (Case B & D), which we considered part of the treatment and operation sub-module. Also, a module was mentioned for advice on behavioral problems (Case B), which we considered part of the Clinical reminders & alerts and referral sub-modules. There were no inconsistencies between the interviewees. Some functional problems were also stated. The system administrator from Case B explained that it was not possible to generate a complete dossier with an overview of the patient status from the HIS. However, the HIS developer mentioned this as one of the strong points of HIS B. From the interviews, sometimes there seems to be a mismatch between the functionalities of the HIS and what is known within the care organization about the HIS. One of the reasons for this mismatch is that often not all modules from the HIS were used or ordered from the software vendor. Another reason for this mismatch was exposed by the system administrator from Case C, who mentioned that care providers sometimes indicate that there is a lot they can do with the HIS and that the overview is sometimes lost. In general, the software developers were more aware of the modules and external entities of the HIS than the system administrators.

Layered view

The layered view for the four case studies is presented in Figure 4.7. When presenting the layered view of [186] to the interviewees, they all agreed that they had a similar layered structure in their HISs. Only as part of the data management layer, the sensor fusion sub-module was only available in cases C and D.

Deployment view

The deployment view is depicted in Figure 5.5. Cases B, C, and D had a similar deployment; they were all structured as a web application running on a server, which the client can reach via an Internet browser. This client could either be a desktop, laptop, tablet, or mobile phone. The HIS from case A could be accessed via a Windows virtual machine (VM) that had the HIS running. The databases from HIS A are hosted at the local GP practices. HISs B and C also have a native Android and iOS app with a simplified version of the HIS. The three HISs working in practice (A, B, C) had some possibility for the client/patient to view their data via an app released in own development (A, B, C) or via a link with a commercial medical application (A).

5.3.3 Big data architecture

As stressed in the Introduction section, a lot of data is collected in the care for people with an ID. This data has considerable potential for research to increase the health and quality of life for this vulnerable population. However, several problems arise when this data needs to become available in a big data analytics system.

Privacy

To get care organisations and their residents to share data, guarantees must be given first to safeguard the privacy of the people whose data is in the system. Of course, a big data analytics system must comply with legislation and follow the associated General Data Protection Regulation (GDPR). Requesting permissions for data sharing is an important aspect of this. However, people with intellectual disabilities cannot always make these kinds of decisions themselves and are, therefore, dependent on their legal representatives. This special situation must be kept in mind when data sharing permissions are requested.

Data in a big data analytics system needs to be pseudonymized before it can be prepared for re-use; this is necessary to safeguard anonymity. No matter how well data is pseudonymized, it can still often be traced back to an individual when looking at small groups of individuals. Analyses done with the big data analytics system will, therefore, always have to be tested for traceability before they are published. Privacy does not only concern people with intellectual disabilities, but also the employees who enter the data and others who are mentioned in the data. The system administrator from Case C mentioned that clients viewing their data via patient portals can lead to problems. Consequently, care providers might sometimes alter their registrations not to cause conflicts with the patients or their representatives. Furthermore, according to Case A, the patient portal also leads patients to view some episodes of their data and request to remove them from the HIS. This can sometimes endanger the continuity of care.

All HISs from the case studies could export data for research purposes. They can all pseudonymize the structured fields but not the textual fields. Patients or clients (or their legal representatives) give permission to use their data when first registered in the HIS. If a researcher wants to research the textual data, explicit informed consent needs to be asked from the study subjects. HISs B and C can register these informed consent forms in their system. In all cases, patients or clients can get their data removed from the HIS or retract their permission to share data, as in accordance with Dutch law.

Security

In the literature, there are three properties of security: Confidentiality, Integrity, and Availability (CIA) [197–199]. Confidentiality aims to prevent unauthorized access to confidential data. Integrity seeks to prevent unauthorized data modification, and availability ensures that data is available to authorized people whenever they need it. There are different tools to assure for each of these properties. To ensure confidentiality, the following tools could/should be used: Encryption, Authentication, Access Control, and/or Authorization. There are two types of tools for integrity to detect whether data has been modified: Prevention mechanisms and detection mechanisms. Prevention mechanisms are needed to avoid corruptible use of the systems; this can be accomplished by tools such as Authentication, Access control, Message signing, and cryptographic techniques. Suppose corruptible use of the system is made despite the prevention mechanisms. In that case, it is important to detect this, which can be done with the detection mechanisms such as intrusion detection and prevention. To assure availability of the system, it is vital to use the following tools: Redundancies, backup, multiple email/Domain Name System (DNS)/ Dynamic Host Configuration Protocol (DHCP) servers, multiple network paths to Internet Service Provider (ISP), firewall, isolation of the organizational internal network from the Internet, allowing some packets to pass, and blocking others. A big data analytics system for ID care should have these security mechanisms in place. Case C showed a potential security issue. When viewing a PDF, the file was automatically downloaded to the local PC, which could lead to data leaks when the local PC is accessed unauthorized. HIS B had a similar issue with .docx and .xlsx files. As this is known, the provider of HIS B strongly advised the end user to encrypt their local disc(s) to mitigate this potential risk.

Big data architecture

Before the big data analytics system is ready to receive data, the privacy issues discussed above must first be safeguarded and the security configuration is set up accordingly. Data in HISs is naturally not available as a public data source. Guarantees must be given to care organizations and their residents to safeguard the privacy of the data. Furthermore, the data in ID care is in multiple HISs and formats. Various systems are used within healthcare institutions, often more than just electronic patient and client dossiers. In addition, dozens of different systems from multiple software vendors are used, resulting in different data formats. A big data analytics system must integrate all these data sources and prepare them for analysis. There is also the volume of the data. Per individual, data is often reported daily. These individuals often live a large part of their lives in a care institution, resulting in thousands of data points per person. With over 90,000 adults with ID living in residential care in the Netherlands [24], this leads to high data velocity.

The big data analytics system architecture for ID care is depicted in Figure 5.6. This architecture is based on the reference architecture described in [191], which is one of the de-facto standards for big data architecture design [99, 200]. We selected the elements from the reference architecture that we considered elemental for developing a big data analytics system for ID care. The big data architecture consists of seven functionalities, from the raw data sources towards interfacing and visualization of the data. We added an extra dimensionality to the reference from [191] in order to visualize how the privacy and security of the system could be safeguarded. Additionally, the entire big data analytics system needs to be surrounded by a firewall to enhance the system's security.

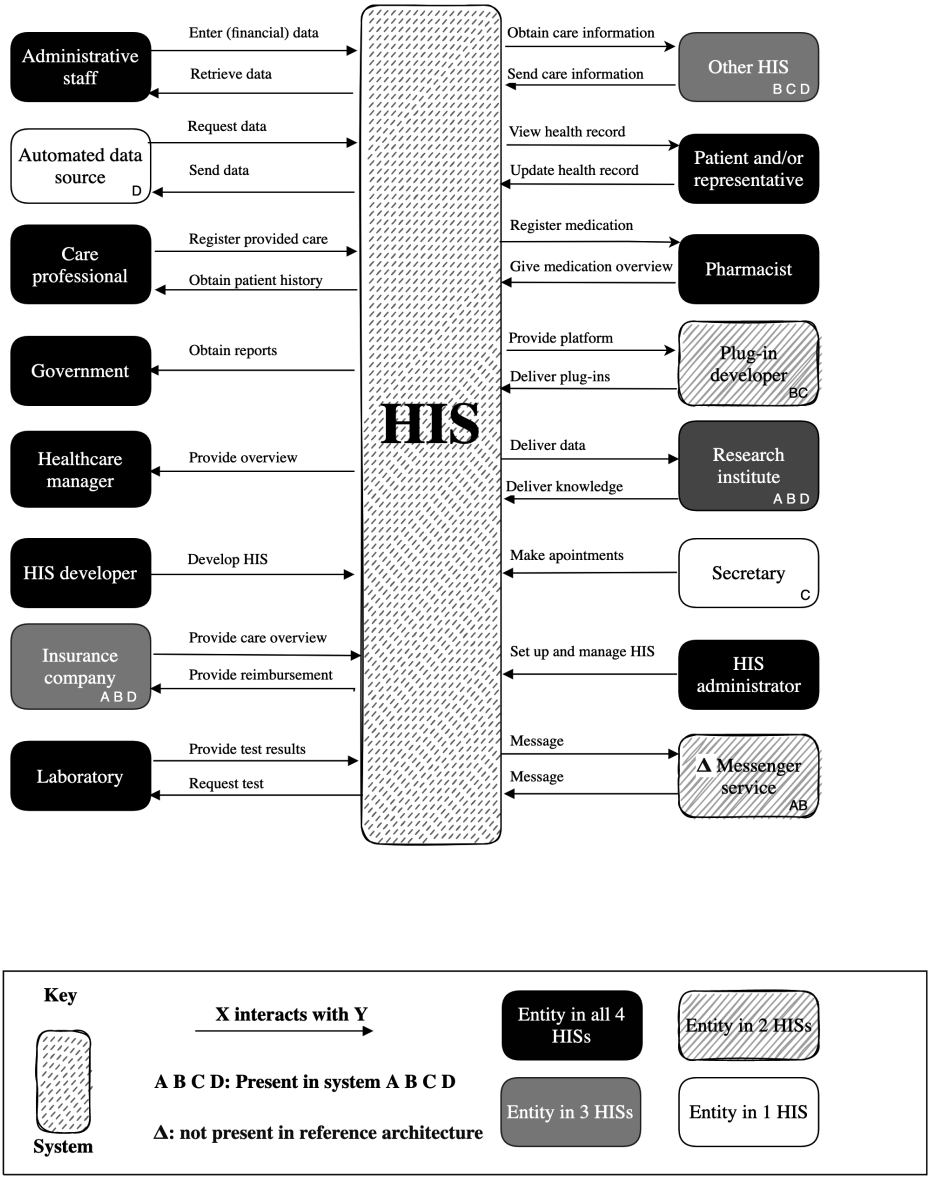


Figure 5.2: Context diagram for the four cases

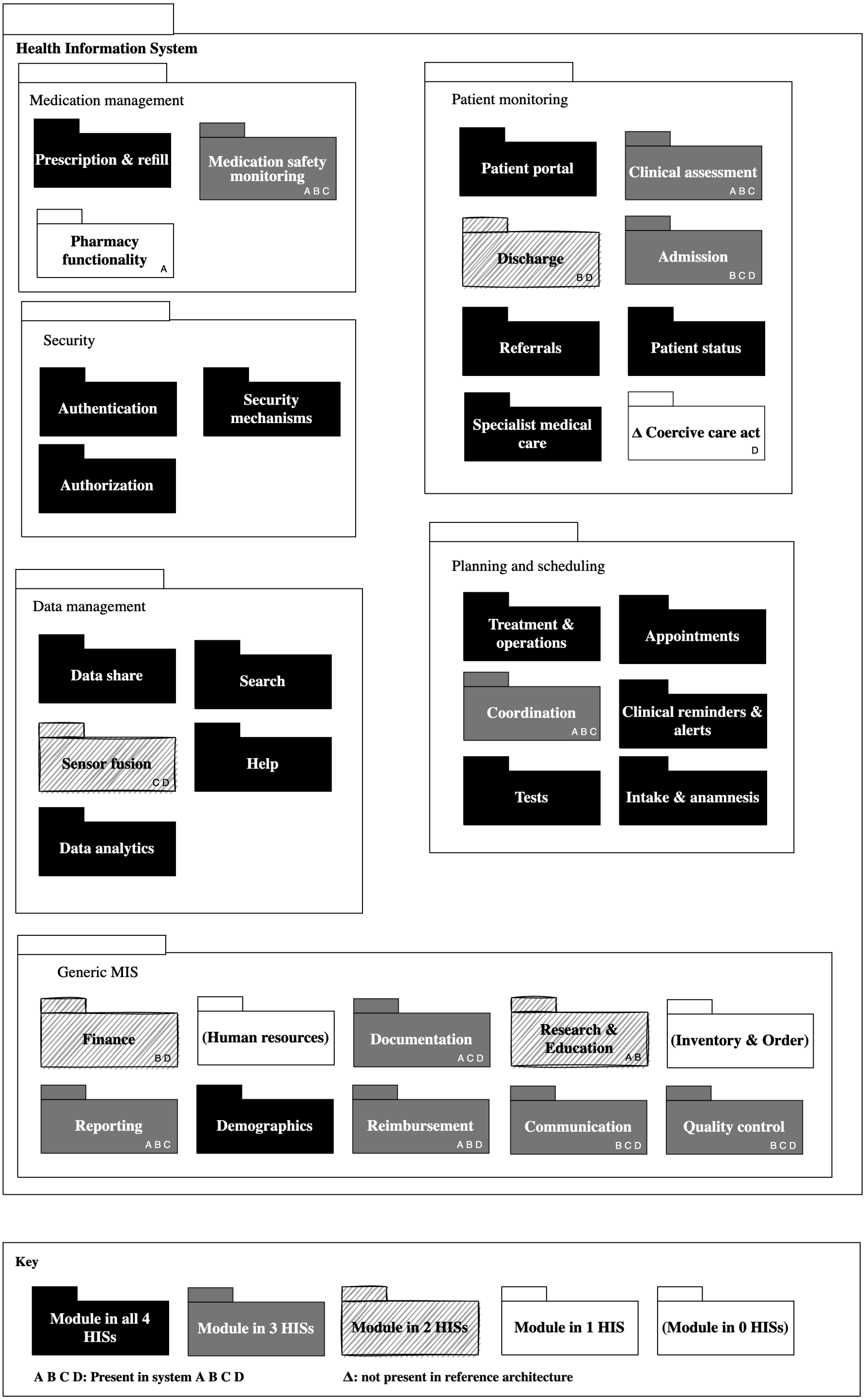


Figure 5.3: Decomposition view for four cases

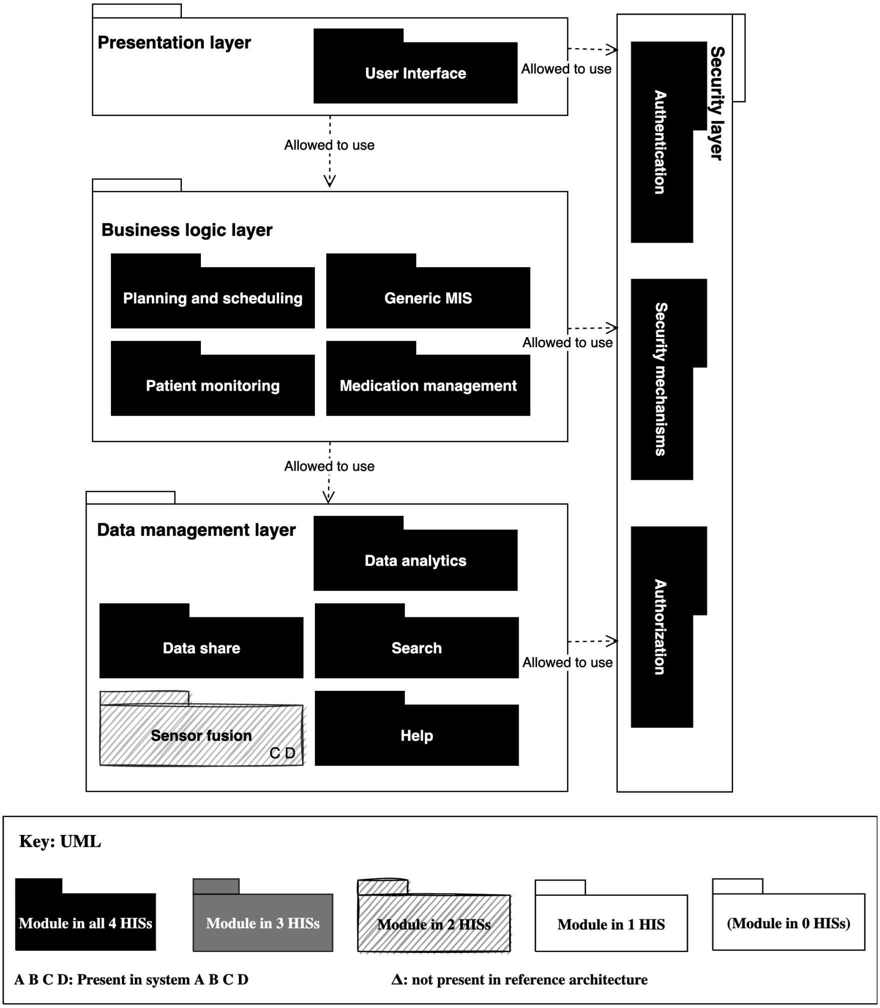


Figure 5.4: Layered view for the four cases

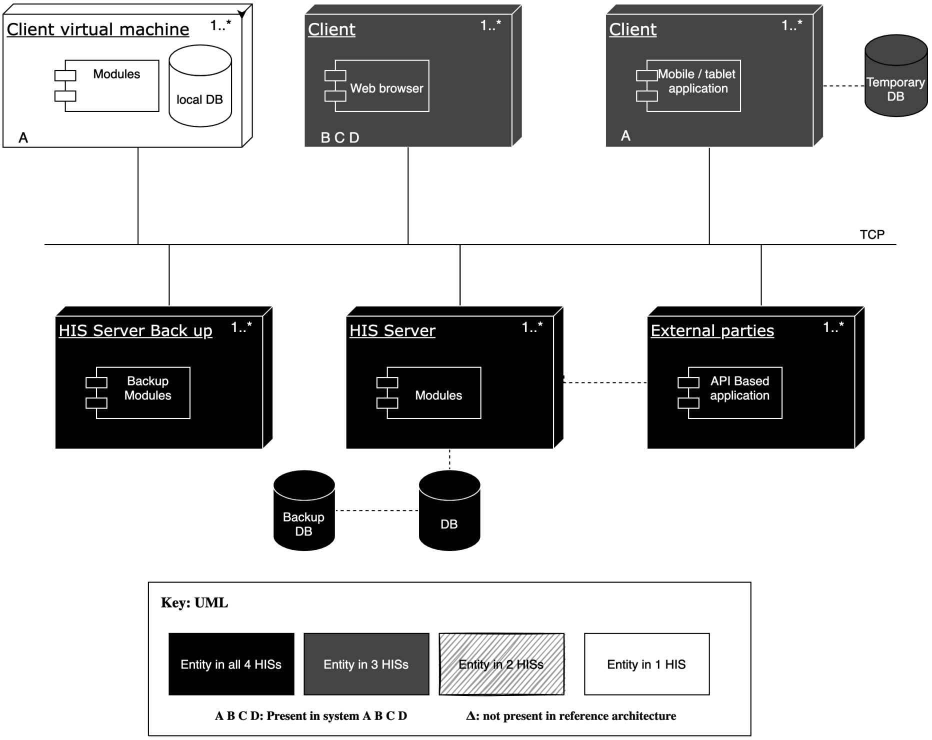


Figure 5.5: The deployment view for the four cases

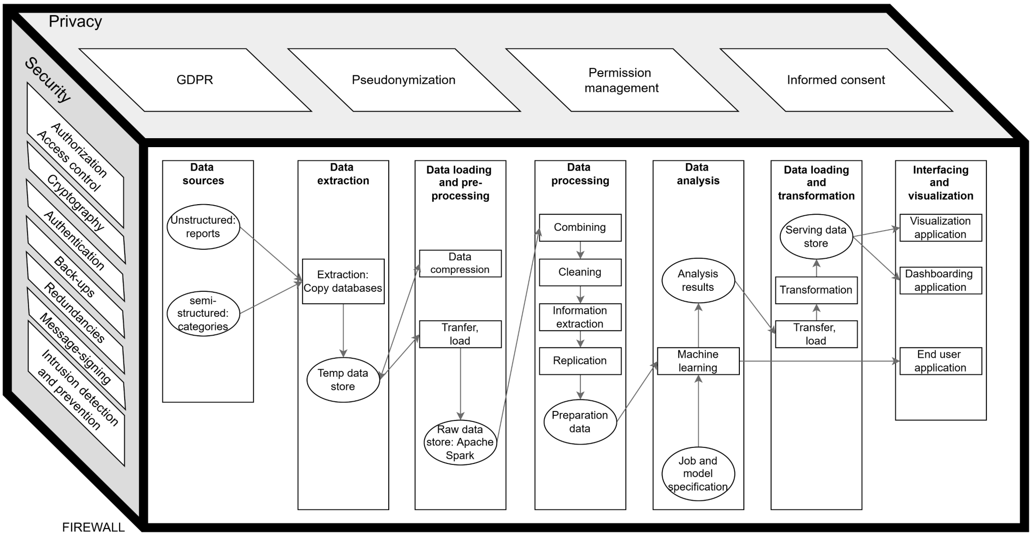


Figure 5.6: General design of big data architecture for ID care, adapted from [191]. Data store is represented as an ellipsis, functionality as a rectangle, and data flow with an arrow.

5.4 Discussion

5.4.1 Case study

The HISs studied have a very different user-interface, and could be accessed in different ways. The HISs contain largely similar entities (context diagram) and 31 out of 35 modules were shared between cases (decomposition view). Differences observed could be largely attributed to the HIS being either an EPD or an ECD. The ECD (Case C) is focused on the continuity of care whereas the EPDs are focused on clinical care which is characterized by episodes. The data from HISs is very privacy sensitive by its nature. All existing HISs studied have built-in methods to export the data anonymously. However, this is limited to (semi-)structured fields: unstructured textual data cannot be exported anonymously. In addition some potential security issues with downloading files were reported. The layered views for the four cases indicated a similar modifiability and portability which facilitates data extraction from the data management layer.

We also noticed that the way of reporting varies from one care facility to another. Each care organization has made its own registration standard, which can complicate analyses. In addition to this not-structured textual registrations, much data is semi-structured or coded (e.g. ICD, CPT, etc.). Data also differs with respect to format e.g. text, documents, images, and videos. A big data analytics system should be able to handle these different data formats.

5.4.2 Big data analytics system architecture

From the interviews we noticed that the majority of the data recorded is in the form of unstructured text. Care providers often spend much time writing reports, e.g. to inform the new shift. A big data analytics system should preferably handle the textual data since this is a central part of the HISs. Therefore, preparations for pseudonymizing textual data should be made prior to analysis. This pseudonymization should include names, addresses, telephone numbers and so on, not only of clients and patients but also of health professionals and others mentioned. Algorithms such as Deduce [201] which was trained on Dutch medical texts, can help in fulfilling this need.

A complicating issue in big data architecture is the necessity to meet all legal requirements for data re-use for research, these legal requirements may differ from country to country. Within the European union, with its GDPR [202], a big data architecture will need to host a facility that allows the exclusion of individuals' data when they or their legal representatives request this. In addition, many Dutch care organizations have an informed consent and an institutional review board in place, also when this is not required according to the Dutch medical research involving human subjects act (WMO [203]).

Although not represented by the big data analytics systems architecture, any use of this system should address the heterogeneity of registrations between and within care organizations and their employees. Also one should realize that only the data that is registered digitally can be analyzed. In the Dutch setting, most data is registered digitally, still data will not be complete, for example day care centers might still register on paper.

5.4.3 Study strengths and limitations

In this study we have adopted a multi-case study on four different HISs in ID care. In further research a broader study could be carried out including other case studies from the 52 distinct Dutch HISs identified elsewhere [182]. However, with this multiple-case study we believe that we have a represented the heterogeneity of HISs used: ECD vs EPD, scale, profession, and time of introduction. This study was limited to Dutch HISs, as the organization of care for people with ID differs largely between countries [63, 204, 205]. HISs suitable for care for people with ID might also be useful for other populations in long term care, such as the elderly, and people who require a long term stay in psychiatric institutions. Indeed, all three existing HISs were also used in elderly care (Case A, B, C). Therefore the presented big data architecture could also be informative to other types of long term care.

Although data sources and extraction may be specific to the Dutch organization of ID care, the data processing, analysis, and visualization are not country-specific. Thanks to the modular set-up of the presented big data architecture, it is not only useful for other countries, but may also contribute to cross-country, international, data systems.

5.5 Conclusion

We provided insight into the HIS architectures from four cases in Dutch ID care in a aggregated manner. Future researchers can use this knowledge to enhance further the HISs used in ID care or to develop new HISs. Based on the commonalities and variabilities of these HIS architectures together with identified privacy and security concerns we provided a big data architecture for Dutch ID care. When built according to this architecture, the big data system will allow for better re-use of the routinely collected data in ID care. Our future work will include developing the big data system according to the provided architecture while satisfying the required privacy and security concerns.

Appendix

Semi-structured interview questions:

System in place

1. Which systems are used at your healthcare organization
2. For what purpose is what system used?
3. What do you consider most important about your system?
4. What measures are in place regarding privacy?
5. What about the interoperability of the system?
6. Is the data in the system suitable to conduct research with?
7. What are the most common problems with your system?
8. What is your system missing?

System under development.

1. How would you describe the system?
2. What is the main reason for developing the system?
3. What do you consider most important to the system?
4. Who are the end users of the system?
5. Can the system be further expanded to other forms of long-term care?
6. What do you think will be the most used functionality of the system?
7. What measures are in place regarding privacy?
8. What about interoperability of the system?
9. Will the data in the system be suitable for doing research with?

Main questions

1. Are you missing any relationships with external entities in the context diagram?
2. Do you find that some relationships in the context diagram are redundant?
3. Create the context diagram based on the method described above
4. 1, 2 but for the decomposition diagram
5. 1, 2 but for the layered diagram
6. 1, 2 but for the deployment diagram
7. Are you missing a diagram?
8. How do you think this reference architecture could be improved?

Data analytics

Study 5: CORD-19 Text mining

6

Background: COVID-19 has been announced as a new coronavirus disease by the World Health Organization (WHO). At the time of writing this article (April 2020), the world is drastically influenced by the COVID-19. Recently the COVID-19 open research dataset (CORD-19) was published. For researchers on intellectual disability such as ourselves, it is of key interest to learn whether this open research dataset may be used to investigate the virus and its consequences for people with an intellectual disability.
Methods: From CORD-19 we identified full text articles containing terms related to the intellectual disability-care and applied a text mining technique, specifically the TF-IDF analysis in combination with K-means clustering.
Results: 259 articles contained one or more of our specified terms related to intellectual disability. We were able to cluster these articles related to intellectual disabilities into five clusters on different topics, namely: Mental health, Viral diseases, Diagnoses & Treatments, Maternal care & Pediatrics, and Genetics.
Conclusion: The CORD-19 open research dataset consists of valuable information about not only COVID-19 disease but also intellectual disabilities and the relationship between them. We suggest researchers investigate Literature-Based Discovery approaches on the CORD-19 and develop a new dataset that addresses the intersection of these two fields for further research.

This chapter is based on an article published in an international journal:

Tummers, J., Catal, C., Tobi, H., Tekinerdogan, B., & Leusink, G. (2020). Coronaviruses and people with intellectual disability: an exploratory data analysis. Journal of Intellectual Disability Research, 64(7), 475-481.

6.1 Background

At the time of writing this article (April 2020), the world that we live in is drastically influenced by the COVID-19, which is a new coronavirus. The first reports on the COVID-19 came in late December 2019 from the Hubei province in China [206]. Since the 11th of March 2020, the COVID-19 is officially a pandemic [207] . At the time of writing this article, there are more than 900,000 confirmed cases and over 45,000 deaths, according to the World Health Organization (WHO) [208].

COVID-19 is spread by human-to-human transmission via small droplets or direct contact [209]. People with an Intellectual Disability might have a higher risk of getting infected by the virus than others for two reasons; they often live close to one another in care facilities, and many rely on physical contact with caregivers for their daily life activities. People of all ages with an intellectual disability have a very high risk of early death due to respiratory infections[210], which makes research into coronaviruses that cause upper-respiratory illnesses (e.g. MERS and SARS)[211] pivotal.

To help mitigate the current public health crisis, the COVID-19 Open Research Data set (CORD-19) has been initiated (CORD-19, 2020). In recent years, the large number of existing scientific publications makes it difficult for researchers to identify relevant information from this huge corpus of papers [212]. This is even more of an issue nowadays, in times of the corona pandemic. The articles in CORD-19 are fragmented, scattered over multiple topics, and thousands of articles are added on a weekly basis. The clustering of articles into groups with similar topics allows to map commonalities and unexplored issues, and helps to efficiently pave the way for further research. Cluster analysis, as used in text mining, helps researchers and other users of these bibliographic databases to get a quick overview of the topics [213]. Because of the clusters, researchers can more effectively decide what topics are sufficiently covered to deserve a systematic review. Furthermore, these clusters may help policy makers and researchers to identify missing topics in research on intellectual disabilities and COVID-19. Currently, it is unknown whether the CORD-19 can be used to investigate the virus and its consequences for people with an intellectual disability. More information on the CORD-19 set can be found in the Methods section.

For researchers on intellectual disabilities, it is of key interest to know what is already in the literature on people with an intellectual disability and coronaviruses. With the help of an exploratory analysis using a text mining technique on the CORD-19 set, we want to investigate whether the CORD-19 may be useful for research on COVID-19 in people with an intellectual disability. We aimed to find clusters of articles that help researchers to identify interesting research topics on coronaviruses and people with an intellectual disability. The presented research and the methods shared, not only yield clusters of articles, but also can help other researchers when they consider pursuing a systematic literature review, and policymakers and researchers to identify research gaps.

The next sections of this paper are organised as follows: The methods section presents the dataset and the tools which were used to analyse this dataset. The paper continues with the results of the exploratory analysis and ends with the discussion.

6.2 Methods

In this section, we present a brief description of the dataset CORD-19, subsequently we explain how we identified articles about people with an intellectual disability and the text mining method that we used in this exploratory analysis.

6.2.1 Dataset

The CORD-19 set has over 44,000 scholarly articles, including over 29,000 with full-text [214]. These articles are all about the COVID-19 and the family of coronaviruses. From the full text articles, 260 contained the word "COVID-19". New articles are added on a weekly basis [215]. All articles were retrieved with a query written in English from the following sources: PubMed [216], a corpus maintained by the WHO [217], bioRxiv [218], and medRxiv [219].For this study, we used the dataset of 23 March 2020. The CORD-19 dataset is put together by the Allen Institute for AI, Chan Zuckerberg Initiative, Georgetown University's Center for Security and Emerging Technology, Microsoft, and the National Library of Medicine (NLM) at the National Institutes of Health. Researchers were asked to apply text and data mining tools on this dataset to develop new insights into the COVID-19 via the Kaggle platform, which is a machine learning and data science community owned by Google Cloud ([220].

6.2.2 Identification of people with an intellectual disability

From the 29,000 articles with a full-text, we aimed to identify a subset of articles that mention people with an intellectual disability. We decided to search for both synonyms of "Intellectual Disability" and specific syndromes. We came up with a list of syndromes based on the website of the Dutch Expertise centre intellectual disabilities [221] and added syndromes based on our own domain knowledge. By combining synonyms of intellectual disability and syndromes, we searched for the following terms in the full text of each article (NB: not case sensitive):
*intellectual disab**, *learning disab**, *mental retardation, cognitive disab**, *mental disab**, *down syndrome, fragile x , prader willi, williams syndrome, fetal alcohol spectrum disorder, rett syndrome, velo-cardio-facial syndrome, angelman syndrome, tuberous sclerosis complex, cornelia de lange syndrome*
The numbers of articles identified with this search are presented in Table [ref] in the Results section.

6.2.3 Data analysis

Text mining is: ". . . . automatically extracting information from different written resources" ([222], 2003, page 1). In our case we started with the TF-IDF algorithm to calculate the importance of each word for each article in the Intellectual Disability subset compared to the importance of that word in the rest of the CORD-19. Then, the outcomes of the TF-IDF were fed into the K-means clustering algorithm to cluster the articles. Articles from different clusters can have some important words in common, therefore there will most probably be some overlap in the top words between the clusters. For this data analysis, we have made a Kaggle notebook, which is available via the following link: `https://www.kaggle.com/joeptummers/covid-2019-id-paper-version`. Python version 3.6.6 [223] was used as the programming language in combination with the pandas package for data analysis and manipulation [224], and, the scikit-learn package for the (machine learning) algorithms [225]. For this exploratory data analysis, we started with the TF-IDF statistical measure (Term Frequency * Inverse Document Frequency), which is widely used in text mining and information retrieval. This statistical measure reflects how important a word is to a document relative to a larger set of documents (i.e., the corpus) [226]. The general formula for the TF-IDF is as follows:

$$w_{i,j} = tf_{i,j} * log(N/df_i) \tag{6.1}$$

With:
$w_{i,j}$ = the TFIDF characteristic for a term i in document j from the subset, that is the articles identified by means of the search
$tf_{i,j}$ = frequency of term i in document j divided by the total words in document j from the subset
N= total number of documents in the corpus
df_i = number of documents in the corpus containing term i

We calculated the TF-IDF characteristics for each term in each article that contained one of the intellectual disability-related terms. We used the complete set of documents (+-29,000) as the corpus to calculate the Inverse Document Frequency. We excluded English stop words (e.g., the, and) and words that appeared in 95% of the articles from the analysis, to speed up calculations. Furthermore, we used the stem of each word to get rid of plurals, tenses, and affixes (e.g., viruses –> virus, infection –> infect, infected –> infect). More parameters can be accessed from our Kaggle notebook. The TF-IDF values were used as input for the K-means algorithm in order to cluster the articles into different clusters. The objective of K-means is to divide n observations into K clusters, in such a way that each observation belongs to the cluster with the nearest mean. K-means does so by minimizing the squared error between the mean of a cluster and the elements in the cluster. The goal of K-mean is to minimise the sum of squared errors over all K clusters [227].

$$O = \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} ||\chi_{k,j,i} - c_k||^2 \tag{6.2}$$

With:
O = The objective function
K = The number of clusters
J = The number of articles
I = The number of terms
$\chi_{k,j,i}$= TF-IDF score for term i in article j in cluster k
c_k = mean TF-IDF for cluster k

Of the several available approaches to determine the number of clusters in K-means clustering [228], we chose to use the widely accepted elbow method. In the elbow method one starts with K = 2, and increases K by 1 in each step and calculates the sum of squared errors (Formula 2) for each step. While increasing K, the sum of squared errors will decrease dramatically in the beginning and will reach a plateau after increasing K further. The K where this happens, is called the elbow and that is where the value for K is set [229].

The TF-IDF analysis yielded one TF-IDF value for each word for each paper (e.g. 259 values for the word "quarantain"). This means that each article is described by about 8.000 variables, and all these variables were used for the K-means clustering. For a visual representation of the 8,000 variables for the articles, we needed to reduce the number of variables by means of a data reduction technique. We chose Principle Component Analysis (PCA) which combined all these variables in two linear combinations (so-called principal components) in a way that explains as much variance in the full data as possible [230]. We chose two principal components to allow for a visual (two-dimensional) representation. In this visual representation every paper is one dot, labelled by the cluster it belongs to.

6.3 Results

After loading and pre-processing the data, we ended up with 26,055 articles that contain the full-text article. In this section, we present the number of articles the terms related to intellectual disabilities appeared in, and continue with the TF-IDF characteristics.

6.3.1 Papers containing intellectual disability care-related words

The number of articles in which the terms related to intellectual disability are present is shown in Table 6.1. Out of the 26,055 articles from the CORD-19 set, a subset of 259 mentioned one of the terms in their full text. Mental retardation was the most identified synonym for intellectual disability, and Down syndrome was the most identified syndrome. Fetal alcohol spectrum disorder and Velo-cardio-facial syndrome were not found in the full texts. The subset of 259 articles was used for the TF-IDF analysis.

Table 6.1: Number of articles with the term present in the full text. Some articles mentioned multiple terms, therefore the total is lower than the sum of the individual terms.

Term	N
Intellectual disability	30
Learning disability	26
Mental retardation	123
Cognitive disability	5
Mental disability	24
Down syndrome	50
Fragile x	33
Prader Willi	1
Williams syndrome	1
Fetal alcohol spectrum disorder	0
Rett syndrome	2
Velo-cardio-facial syndrome	0
Angelman syndrome	2
Tuberous sclerosis complex	9
Cornelia de Lange syndrome	1
Total	259

6.3.2 TF-IDF and K-means

The elbow of the plot of the sum of squared errors versus K appeared to be at K = 5 (Figure 6.1). Therefore, we decided to use K = 5 and let the algorithm group the articles in five clusters. These clusters are depicted in the Principal Component plot in Figure 6.2. An overview of the subset of 259 articles and the clusters they belong to can be found at the end of the Kaggle notebook. Table 6.2 shows the five clusters in the subset, with their top 10 TF-IDF terms. These top 10 words are stemmed; therefore, they might look a bit unusual. We also added a column with "topic" that contains our own interpretation of the top 10 words. The biggest cluster was the "Genetics" cluster, with 105 articles, and the smallest cluster was the "Mental health" cluster, with 26 articles.

Table 6.2: Five clusters, their top 10 stemmed words from the TF-IDF analysis on the subset, their topics and number of articles in the cluster.

	Top 10 words based on TF-IDF score	Topic	N
1	disord, social, diseas, studi, public, mental, patient, psychiatr, ptsd, health	Mental health	26
2	viral, may, caus, patient, diseas, immun, cell, vaccin, virus, infect	Viral diseases	38
3	cell, treatment, disord, studi, infect, diseas, zikv, may, children, patient	Treatments	57
4	breastfeed, hospit, patient, pneumonia, vaccin, infect, bronchiol, rsv, infant, children	Maternal care & pediatrics	33
5	virus, interact, dna, bind, express, activ, rna, gene, cell, protein	Genetics	105

Footnote: Using a more extensive set of stopwords, might have filtered "may" and "caus" out. "zikv" is the often used abbreviation for Zika Virus.

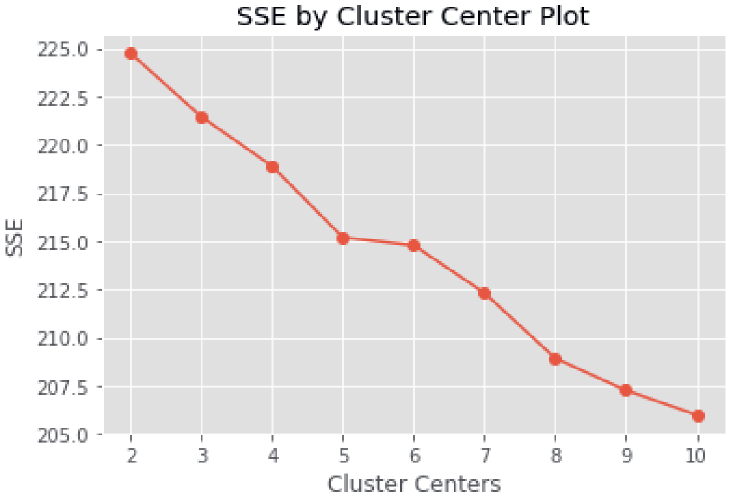


Figure 6.1: The sum of squared errors versus the number of clusters (See formula 2). The "elbow" of the graph appears to be at K=5..

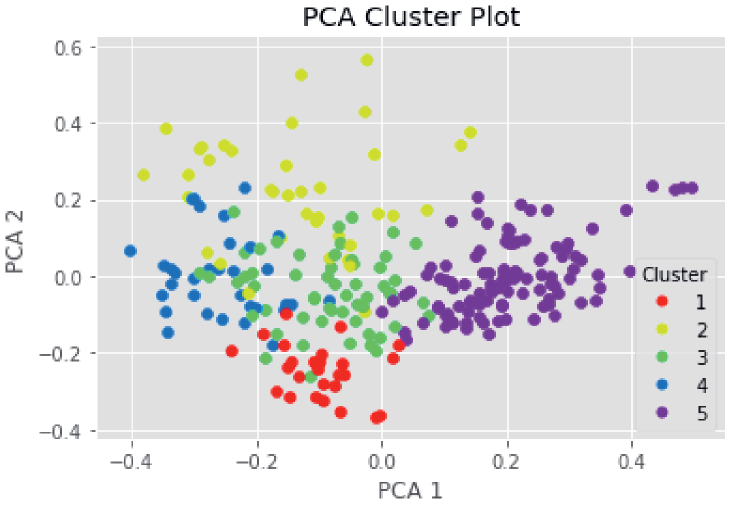


Figure 6.2: PCA plot with two principal components. Contains 259 data points (articles), divided over five clusters.

6.4 Discussion

From the CORD-19 set, we identified that 259 out of the 26,055 articles had one of the terms related to the intellectual disability-care in their full text. We are aware that this list of used terms is not exhaustive, but considered the list sufficient for exploring whether CORD-19 has any potential for research on people with an intellectual disability. We did not include the 22Q11 deletion syndrome, and Autism Spectrum Disorders on purpose since these syndromes do not always co-occur with an intellectual disability. We do not expect a big rise in the number of articles identified if we had added more terms. The current number of 259 articles seems to be sufficient for a systematic literature review, particularly one aiming at genetics aspects. Moreover, thousands of articles are added to the CORD-19 dataset every week. Please note that our method and code are publicly available, and can be re-used by other researchers within just a few minutes.

Due to the application of the K-means algorithm on the outcomes of the TF-IDF calculations, we were able to cluster the articles that contained the intellectual disability-terms in five clusters on different topics. Researchers on intellectual disabilities can use these clusters to identify research areas on the topic of Coronaviruses and people with an intellectual disability, and, possible relevant articles already published on it. However, we have to be aware that setting the number of clusters at five was, to some extent, a design choice. When fixing the number of clusters at four, we saw more overlap of the clusters in the PCA plot, probably because the articles in our cluster number three (from Table 6.2) were dispersed over the other clusters. When we set the number of clusters at six, the extra cluster was a small cluster, very similar to our cluster five (Genetics). Repeating the analyses on an extended edition of CORD-19, may result in a different number of clusters with other topics.

Currently there are search engines available for the CORD-19 [231] that can be useful to identify papers related to a certain topic. However, these search engines do not map the commonalities between different articles nor do they help to identify what has not been investigated yet, and this is where the methods described in the present paper have their added value.

Our exploratory analysis has shown that the CORD-19 set can be useful for researchers on intellectual disabilities and may be approached with text mining techniques. If we aim to create new hypotheses on the relations between coronaviruses and people with an intellectual disability based on literature, we need to utilise other text mining-based techniques. One approach that seems to have great potential is the Literature-Based Discovery (LBD). This approach uses various computational methods to discover previously unknown links between two pieces of existing knowledge by analysing their relevant pieces of literature [232, 233]. In our case, one piece of knowledge would be the COVID-19 Open Research dataset and the other a set of articles on people with an intellectual disability. Especially, the cluster similarity technique [234] and bibliographic coupling technique [235] seem to be suitable to construct new hypotheses on possible relations between corona viruses and people with an intellectual disability.

Concluding, the CORD-19 has shown to be interesting for research on intellectual disabilities and coronaviruses. At first sight, there seemed to be relatively few articles (259 out of 26,055) in the COVID-19 Open Research Dataset related to intellectual disabilities. Nonetheless the full text articles were approachable by means of text mining techniques. With TF-IDF and K-means clustering we were able to identify five clusters of articles on different topics. Researchers on intellectual disabilities can use the presented five clusters to identify research areas on intellectual disabilities and coronaviruses and to decide on pursuing a systematic literature review (or not). Also, they may apply our approach again and again on the weekly updates of CORD-19, as this will take little time. For creating new hypotheses on the relations between coronaviruses and people with an intellectual disability, based on CORD-19, we suggested other text mining approaches.

General discussion

General discussion 7

7.1 Objective

The care of people with ID involves a large number of care providers and HISs resulting in an enormous amount of data entered in a range of HISs by different people daily. There is an increasing interest in the re-use of data as collected in HISs on people with ID to improve care provided, to gain knowledge on the health of people with ID, and to deal with the under-representation of people with ID in general medical research. The objective was to develop a framework for a big data system that combines data from multiple HISs and facilitates secondary analysis.

7.2 The status quo

In the care for people with ID, many different HISs are used (Chapter 2), each containing a wide variety of data types. Considering the care organizations in the Netherlands, there are usually two types of systems: the electronic client dossier (ECD) and the electronic patient dossier (EPD). The ECD systems and the the range of data types in them are relatively similar, although there are many ECD vendors. ECDs contain mainly texts in the form of daily registrations and periodic questionnaires to ensure continuity of care. Although these daily registrations and texts are stored in each ECD in a slightly different format, the content across systems is similar.

By their nature, EPD systems are somewhat more structured than ECDs because of their origins as GP or hospital information systems where much is encoded and stored as semi-structured texts. The variety of the EPDs used in ID care is large, and systems are less similar than ECDs. Compared with ECDs, more research has already been done into EPDs regarding how the data can be analyzed when brought together [236, 237]. A big data system for the ID care sector may take advantage of these developments and use a certain structure to combine data from the Netherlands's most frequently used ECDs and EPDs.

A set of more "soft" problems were encountered in this thesis, influencing the development of a big data system. There are plenty of opportunities for re-using data in the care for people with ID. An important element is arranging informed consent to participate in research [238, 239]. People with ID often depend on their legal representatives to permit them to participate in research. Furthermore, care organizations often have to deal with a jumble of consents for research and data sharing, creating an organizational challenge to keep this streamlined. Not all HISs in the care for the intellectually disabled have an equally good possibility for keeping track of these consents. A big data system could play a role in keeping track of consents. When a researcher wants to research a certain registration type, the big data system could facilitate obtaining the necessary permissions from a medical ethics review committee, care institutions, and clients or legal representatives. Permission management could be done through a central location where the consents are kept, and the legal obligation to revoke consents can be met. Only when all permissions are in, the data can be delivered to the researcher in his/her secure environment where the dataset can be analyzed.

During the different aspects of this research, a vast set of technical problems were encountered that need to be solved before a big data system can be developed. Data from HISs is already used in research on people with an ID, usually this research is conducted with aggregated data such as reimbursements [20, 240, 241]. However, there is still a large potential of rich, raw data available which is not yet used in research. Compared to short-term care, such as GP or hospital care, a big data system for the ID sector has the great advantage of containing data on the entire life of an individual with an ID. For the normal population, data is usually only available on being ill. When ill, a person comes in contact with a GP information system and possibly a hospital information system, where the illness is registered. In the care for people with ID, this information is available, but also data prior to illness is available. If the data from these systems can be linked in a big data system, a great research potential will emerge.

In order to use this unexploited source of (textual) data for research, it is necessary to use artificial intelligence techniques. AI techniques can be used to establish links between cause and effect. AI has great potential for preventive healthcare for this vulnerable population, and preventive, AI based health care could lead to better and cheaper care for people with ID. Text mining and natural language processing are fields of AI that can reveal connections in texts [242]. Until recently, the algorithms behind these techniques were mainly trained on English texts. Due to the enormous rise in research on AI algorithms, more of these techniques are also available for the Dutch language [243, 244]. The Veracity of data may affect the result of the analysis done. The question is what is actually registered in healthcare and what the care professionals want to be shown. The interpretation of the healthcare provider registering also influences the registration; for example, someone may think that a person with an ID can be in pain or stress because of his/her behavior. With the introduction of the client portal, where clients have insight into their files, there seems to be a tendency (consciously or unconsciously) that the care provider does not register everything. Therefore, what is registered is often not something that represents reality.

The long-term care sector in Europe faces a problem with high staff turnover [245]. Experienced personnel can often tell from small changes in the behavior of a person with ID what is going on with him/her, while new and inexperienced staff are less able to do this. With the help of a big data system, data can be easily extracted from HISs, analyzed by researchers in order to find relations between changes in behavior and effects, which can be implemented in existing communication systems. These implementations could, for example, result in a warning for aggressive behaviour for a client when a certain text is entered in his/her dossier. This way, the HIS could support the caregiver even better in making decisions around the client.

It cannot be stressed enough how important it is to safeguard the privacy of the people whose data is in the systems. Privacy concerns not only the people with ID and their relatives but also the caregivers. A big data system must make sure that the data cannot easily be traced back to one person. With pseudonymization, a person can be followed through time while assuring anonymity [201]. However, there remains the risk that data may be traced to a person, especially if it concerns a target group where certain diseases and conditions only occur a few times in the entire population. Non-disclosure agreements for researchers and adherence to the ethical code for handling medical data should reduce this risk as much as possible. Furthermore, agreements will have to be made regarding data publication on small groups.

7.3 Reflection on this thesis

In comparison with current literature, which most often focuses on combing data of one type (e.g. codes, texts) or source (e.g. Radiology [246] or E-health [237]), this thesis focuses on combining data from many different data types from many sources. An interdisciplinary team approached the thesis from multiple (computer science, medical science, and methodology) perspectives, giving a comprehensive view of the subject. This thesis has its limitations; for example, it is focused on the Dutch organization of ID care. Although the national focus, this does not mean the work presented in this thesis is not interesting for others. The results may also be transferred to other forms of long-term care, where data registration is similar. Although the organization of care is different between countries [63, 204, 205], the generic ideas about merging data from different HISs can also be applied there. From the beginning of this thesis, the complete care landscape of Dutch ID care was included to incorporate the wide variety of caregivers involved, and HISs used.

This thesis uses standards to ensure that work can be repeated, supplemented, and changed as needed. This flexibility is reflected in the software architecture chapters (4 and 5), where there is freedom for customization, addition, and reduction. Additionally, using standards provides unity of language, which is very important in interdisciplinary research, such as this thesis [247]. Suppose new standards are developed for software architecture design, such as a new Unified Modelling Language (UML) version. In that case, these standards need to be implemented in work presented in this thesis to improve the comprehensibility of the presented architectures.

Based on the knowledge gained about the ID care sector, the HISs used, and the problems involved in re-using data, POS^T4V is proposed for a big data system in ID care.

This abbreviation stands for: **P**rivacy, **O**rganisation of care, **S**ecurity, enabled by **T**echnical solutions to be able to act with **V**olume, **V**elocity, **V**ariety, **V**eracity of data in ID care.

When developing a big data system, privacy and security are of uttermost importance to protect the vulnerable population from improper re-use of their data. The Organisation of care is unique in ID care, involving many different care providers and HISs. To re-use the big data (4Vs) collected in ID care, the Privacy, Organisation of care, and Security need to be covered by technical solutions; therefore the "t" is super-scripted.

7.4 Where to go, and how to get there?

Many interesting (technical) solutions are available to mitigate the discussed problems. Ensuring privacy is key for ID care when applying these solutions. When one starts to apply AI to data from this vulnerable population, extra caution is needed. Outside of ID care there has been an increasing number of harm because of the miss-use of AI technology (e.g. facial recognition surveillance, voter manipulation, mass data collection without consent), or because of design flaws (e.g. loan rejection based on race, medical misdiagnoses)[248, 249]. Therefore, AI ethical guidelines must be followed closely before the results could influence decision-making [250].

"Soft" problems for re-using routinely collected data include subjects such as privacy, security, and informed consents cannot be ignored when developing a big data system to combine data from this vulnerable population. For example, it is impossible to calibrate caregivers, resulting in an heterogeneous dataset of which the quality can be questioned. Using technical solutions (AI, text mining) to solve these problems seems worthwhile. Usually, the principle 'garbage in is garbage out' holds when doing AI [251]. But, looking at emerging techniques like deep learning, the question is whether this statement is true. Deep learning might recognize links in the text in such a way that it also recognizes poor quality registrations and includes them in the analyses labeled as poor registrations. Furthermore, deep learning could also remove the need for manual feature engineering, simplifying the whole process [252].

Registrations in ID care lead to multiple enormous data sets primarily consisting of texts. With the recent increase in processing power of computers and the possibility of cloud computing [253], much more is possible when analyzing large amounts of text. It should be noted that doing these extensive analyses is not a purely positive thing. The computing power required results in large servers that run for hours, resulting in high energy consumption and the associated environmental impact [254]. For a big data system in the

ID sector, it will be necessary to look at how these effects can be prevented or compensated.

Security of the big data system must ensure that no unauthorized access is possible. For security, many precautions must be taken because of the privacy sensitivity of the data collected in the system. One may learn from systems such as the Digital Research Environment [255] that offer cloud computing technology in the most secure way possible. For example, no data can be taken from the server without first requesting permission. On the other hand, security should not hinder the data analyses, and the use of the previously mentioned techniques should be possible when acting in accordance with ethical guidelines. Because of the amount of data, cloud computing technologies from third parties will have to be used. These parties must be certified according to the relevant laws and regulations.

In the Netherlands, professional associations are an important link in ID care. They often contribute to the guidelines that care professionals follow to do their jobs well. The professional associations need to realize the big data system's potential and contribute to the system's development. Finally, there are the software vendors of the electronic patient and client dossiers. They will have to build functionality into their software packages to facilitate data exchange with the big data system in an automated manner. The big data system will be aligned with the software vendor's data exchange requirements, so there are as few barriers as possible to the data exchange. Software providers should work with healthcare organizations to ensure that the data transition is as seamless as possible, with the explicit condition that the permissions of the necessary parties are in place.

There are technical solutions for many of the discussed problems, but for some, the consequences are yet unknown. For these reasons, caution is advised in the developmental process. Starting the development of the big data system will undoubtedly result in new research questions and design challenges which will, in the end, only improve the system.

7.5 Future research

To start building a big data system, getting the main stakeholders on board regarding the developmental strategy is vital. Government organizations have both a funding and a legislative role. Furthermore, they benefit from the results if they can make care more efficient. Non-medical care organization staff such as administrators and directors should also participate in developing a big data system; they have the data and can benefit from the research results on this dataset, which they can implement in their care organization. Researchers are the party who are important for the successful development of a big data system for the care of people with ID because they are the group of stakeholders driving the system's development. Researchers are also the people who can make the big data system a success. If sound research is done on the data, this will positively affect the system's impact. If a complex and flexible big data system is developed that is not used by any researcher, the system will not be a success.

I suggest to start developing a big data system to combine data in Dutch intellectual disability care, on a small scale, very carefully! To be cautious enough, initially more research is needed into the legal aspects of re-using data in ID care. This includes research on what is necessary regarding obtaining informed consent from both the client and the caregiver. A small-scale system should contain data from both ECDs and EPDs from at least two care organizations. This data will not be automatically extracted from the different systems but will be requested in batches from the system administrators in the care organizations. Future research is needed on how the data can be merged in an automated manner. Based on this thesis, it is suggested to start with the daily and incident reports from the ECDs and the textual reports and medical codes from the EPDs, since these data types are always present. The exploratory and design work will have to be done by a software architect, who writes specifications such that a software developer can start building it.

In addition to solving practical problems that such a small system helps to identify, other issues must also be solved. Research is needed for the inclusive development and design of a big data system. Not only people with an intellectual disability and their legal representatives need to be included in the design process, also the caregiver, entering the data in the HISs needs to be included.

Next to the developments on a system level, research on technical aspects is also needed. For example, algorithms currently available in English or Spanish should be made available for Dutch texts. Moreover, it will be necessary to examine whether existing algorithms can deal with current textual data, which contains, for example, abbreviations, sarcasm and double negatives. This examination could be done by looking for the known way to validate algorithms. For example, by researching if it can be extracted from the analyses whether a link can be made between down syndrome and obesity [256]. After the validation of the algorithms, conformation studies should be done on the signals that are picked up with the analyses.

7.6 Take home message

Based on this thesis, it was identified that the success of the development of a big data system for ID care depends on taking POS^T4V into account. When the knowledge is available on how one can merge and re-use these data, the parties involved will have to discuss how this can be done in an automated way.

In short, there is a lot of potential for developing a big data system to re-use data in the ID sector. To start the development everybody must be on the same page: Governments, care organizations, researchers, professional associations, and software vendors. Only together the care of people with ID can be improved.

Bibliography

Here are the references in citation order.

[1] M. M. Yusof et al. 'An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit)'. In: *International journal of medical informatics* 77.6 (2008), pp. 386–398 (cited on pages 1, 12, 30).

[2] F. Folkvord et al. *Benchmarking deployment of eHealth among general practitioners (2018) : final report*. Publications Office of the European Union, 2019 (cited on page 1).

[3] F. F. Odekunle, R. O. Odekunle, and S. Shankar. 'Why sub-Saharan Africa lags in electronic health record adoption and possible strategies to increase its adoption in this region'. In: *International journal of health sciences* 11.4 (2017), p. 59 (cited on page 1).

[4] C. S. Kruse et al. 'Impact of electronic health records on long-term care facilities: systematic review'. In: *JMIR medical informatics* 5.3 (2017) (cited on page 1).

[5] American Psychiatrics Association. *What is intellectual disability*. URL: `https://www.psychiatry.org/patients-families/intellectual-disability/what-is-intellectual-disabilit` (visited on 07/07/2022) (cited on pages 1, 2).

[6] American Association on Intellectial and Developmental Disabilities. *Intellectual disability definition*. URL: `https://www.aaidd.org/intellectual-disability/definition` (visited on 07/27/2022) (cited on page 2).

[7] I. Woittiez et al. 'Zorg beter begrepen'. In: *Verklaringen voor de groeiende vraag naar zorg voor mensen met een verstandelijke beperking. Den Haag: SCP* (2014) (cited on pages 2, 12).

[8] A. A. Beaujean et al. 'A misuse of IQ scores: Using the dual discrepancy/consistency model for identifying specific learning disabilities'. In: *Journal of Intelligence* 6.3 (2018), p. 36 (cited on page 2).

[9] Merriam-Webster. *Intellectual Disability*. URL: `https://www.merriam-webster.com/dictionary/intellectual%5C%20disability` (visited on 07/27/2022) (cited on page 2).

[10] A. M. W. Coppus. 'People with intellectual disability: what do we know about adulthood and life expectancy?' In: *Developmental disabilities research reviews* 18.1 (2013), pp. 6–16 (cited on page 2).

[11] M. Hanekamp et al. *Administratieve belasting langdurige zorg 2019*. Tech. rep. Berenschot, 2019, pp. 1–59 (cited on pages 2, 12).

[12] E. Anskär et al. 'Time utilization and perceived psychosocial work environment among staff in Swedish primary care settings'. In: *BMC health services research* 18.1 (2018), pp. 1–12 (cited on page 2).

[13] C. Sinsky et al. 'Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties'. In: *Annals of internal medicine* 165.11 (2016), pp. 753–760 (cited on page 2).

[14] C. K. Reddy and C. C. Aggarwal. *Healthcare data analytics*. Vol. 36. CRC Press, 2015 (cited on page 2).

[15] C. Lee et al. 'Big healthcare data analytics: Challenges and applications'. In: *Handbook of large-scale distributed computing in smart healthcare*. Springer, 2017, pp. 11–41 (cited on page 2).

[16] A. Kuspinar et al. 'Development and validation of an algorithm to assess risk of first-time falling among home care clients'. In: *BMC geriatrics* 19.1 (2019), pp. 1–8 (cited on page 2).

[17] Y. Lo et al. 'Using machine learning on home health care assessments to predict fall risk'. In: *MEDINFO 2019: Health and Wellbeing e-Networks for All*. IOS Press, 2019, pp. 684–688 (cited on page 2).

[18] P. Melillo et al. 'Identifying fallers among ophthalmic patients using classification tree methodology'. In: *PloS one* 12.3 (2017), e0174083 (cited on page 2).

[19] K. Walters et al. 'Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data'. In: *BMC medicine* 14.1 (2016), pp. 1–12 (cited on page 2).

[20] M. Cuypers et al. 'Cancer-related mortality among people with intellectual disabilities: A nationwide population-based cohort study'. In: *Cancer* 128.6 (2022), pp. 1267–1274 (cited on pages 2, 92, 128).

[21] C. M. Bertoncelli et al. 'Identifying factors associated with severe intellectual disabilities in teenagers with cerebral palsy using a predictive learning model'. In: *Journal of Child Neurology* 34.4 (2019), pp. 221–229 (cited on pages 2, 92).

[22] R. Shankar et al. 'Under representation of people with epilepsy and intellectual disability in research'. In: *PLoS One* 13.6 (2018), e0198261 (cited on page 3).

[23] I. Anagnostopoulos, S. Zeadally, and E. Exposito. 'Handling big data: research challenges and future directions'. In: *The Journal of Supercomputing* 72.4 (2016), pp. 1494–1516 (cited on page 3).

[24] M. Cuypers et al. 'Linking national public services data to estimate the prevalence of intellectual disabilities in The Netherlands: results from an explorative population-based study'. In: *Public Health* 195 (2021), pp. 83–88 (cited on pages 3, 92, 102).

[25] J. A. Hirsch et al. 'ICD-10: history and context'. In: *American Journal of Neuroradiology* 37.4 (2016), pp. 596–599 (cited on page 3).

[26] International Council of Nurses. *about-icnp*. URL: `https://www.icn.ch/what-we-do/projects/ehealth-icnptm/about-icnp` (visited on 07/27/2022) (cited on page 3).

[27] G. Nahler. 'Anatomical therapeutic chemical classification system (ATC)'. In: *Dictionary of Pharmaceutical Medicine*. Springer, 2009, pp. 8–8 (cited on page 3).

[28] M. H. Fisher et al. 'Victimisation and social vulnerability of adults with intellectual disability: A review of research extending beyond Wilson and Brewer'. In: *Australian Psychologist* 51.2 (2016), pp. 114–127 (cited on page 4).

[29] H. Prosser and J. Bromley. 'Interviewing people with intellectual disabilities'. In: *Clinical psychology and people with intellectual disabilities* (1998), pp. 99–113 (cited on page 4).

[30] F. J. Symons, S. K. Shinde, and E. Gilles. 'Perspectives on pain and intellectual disability'. In: *Journal of Intellectual Disability Research* 52.4 (2008), pp. 275–286 (cited on page 4).

[31] J. J. Polder et al. 'Healthcare costs of intellectual disability in the Netherlands: a cost-of-illness perspective'. In: *Journal of Intellectual Disability Research* 46.2 (2002), pp. 168–178 (cited on pages 4, 12).

[32] Object Management Group. *Business Process Model & Notation™ (BPMN™)*. URL: `https://www.omg.org/bpmn/` (cited on page 5).

[33] L. Bass, P. Clements, and R. Kazman. *Software Architecture in Practice*. Addison-Wesley Professional, 2003 (cited on pages 6, 68).

[34] R. Haux et al. *Strategic information management in hospitals: an introduction to hospital information systems*. Springer Science & Business Media, 2004 (cited on pages 12, 30).

[35] R. Haux. 'Health information systems–past, present, future'. In: *International journal of medical informatics* 75.3-4 (2006), pp. 268–281 (cited on pages 12, 30).

[36] World Health Organization. *Everybody's business: strengthening health systems to improve health outcomes. WHO's Framework for Action. Geneva: World Health Organization; 2007*. Tech. rep. World Health Organization, 2007 (cited on pages 12, 30).

[37] R. Skolnik. *Global health 101*. Jones & Bartlett Publishers, 2015 (cited on pages 12, 30).

[38] A.S. McAlearney et al. 'Perceived efficiency impacts following electronic health record implementation: an exploratory study of an urban community health center network'. In: *International journal of medical informatics* 79.12 (2010), pp. 807–816 (cited on page 12).

[39] J.H. Sinard et al. 'Stand-alone laboratory information systems versus laboratory modules incorporated in the electronic health record'. In: *Archives of Pathology and Laboratory Medicine* 139.3 (2015), pp. 311–318 (cited on page 12).

[40] L. Raymond, G. Paré, and M. Marchand. 'Extended use of electronic health records by primary care physicians: Does the electronic health record artefact matter?' In: *Health informatics journal* 25.1 (2019), pp. 71–82 (cited on pages 12, 21).

[41] M. Trudel et al. 'Ceiling effect in EMR system assimilation: a multiple case study in primary care family practices'. In: *BMC medical informatics and decision making* 17.1 (2017), p. 46 (cited on page 12).

[42] D. J. Brailer. 'Interoperability: The Key To The Future Health Care System: Interoperability will bind together a wide network of real-time, life-critical data that not only transform but become health care.' In: *Health affairs* 24.Suppl1 (2005), W5–19 (cited on page 12).

[43] S. V. B. Jardim. 'The electronic health record and its contribution to healthcare information systems interoperability'. In: *Procedia technology* 9 (2013), pp. 940–948 (cited on page 12).

[44] E. J. S. Hovenga. 'Importance of achieving semantic interoperability for national health information systems'. In: *Texto & Contexto-Enfermagem* 17.1 (2008), pp. 158–167 (cited on page 12).

[45] J. Adler-Milstein et al. 'Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide'. In: *Journal of the American Medical Informatics Association* 24.6 (2017), pp. 1142–1148 (cited on pages 12, 21, 68).

[46] F. Alsohime et al. 'Satisfaction and perceived usefulness with newly-implemented Electronic Health Records System among pediatricians at a university hospital'. In: *Computer methods and programs in biomedicine* 169 (2019), pp. 51–57 (cited on pages 12, 21).

[47] J. Kaipio et al. 'Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland'. In: *International journal of medical informatics* 97 (2017), pp. 266–281 (cited on pages 12, 21, 68).

[48] A. A. Holanda et al. 'Use and satisfaction with electronic health record by primary care physicians in a health district in Brazil'. In: *Journal of medical systems* 36.5 (2012), pp. 3141–3149 (cited on pages 12, 21).

[49] K. Soderberg et al. 'Health Reform in Minnesota: An Analysis of Complementary Initiatives Implementing Electronic Health Record Technology and Care Coordination'. In: *Online journal of public health informatics* 8.3 (2016) (cited on pages 12, 21).

[50] G. B. Willis. *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications, 2004 (cited on page 13).

[51] Qualtrics. *Qualtrics survey software*. 2020 (cited on page 14).

[52] M. E. May and C. H. Kennedy. 'Health and problem behavior among people with intellectual disabilities'. In: *Behavior Analysis in Practice* 3.2 (2010), pp. 4–12 (cited on page 21).

[53] S. Kim et al. 'Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital'. In: *BMC medical informatics and decision making* 16.1 (2016), p. 12 (cited on page 21).

[54] G. Marca et al. 'The use of electronic health records in Spanish hospitals'. In: *Health Information Management Journal* 43.3 (2014), pp. 37–44 (cited on page 21).

[55] H. E. Resnick and M. Alwan. 'Use of health information technology in home health and hospice agencies: United States, 2007'. In: *Journal of the American Medical Informatics Association* 17.4 (2010), pp. 389–395 (cited on page 21).

[56] E. R. Pfoh et al. 'Satisfaction after the transition between electronic health record systems at six ambulatory practices'. In: *Journal of evaluation in clinical practice* 18.6 (2012), pp. 1133–1139 (cited on page 21).

[57] J. Viitanen et al. 'National questionnaire study on clinical ICT systems proofs: physicians suffer from poor usability'. In: *International journal of medical informatics* 80.10 (2011), pp. 708–725 (cited on page 21).

[58] M. Topaz et al. 'Nurse informaticians report low satisfaction and multi-level concerns with electronic health records: results from an international survey'. In: *AMIA Annual Symposium Proceedings*. Vol. 2016. American Medical Informatics Association. 2016 (cited on page 21).

[59] B. W. Pickering et al. 'The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: a pilot step-wedge cluster randomized trial'. In: *International journal of medical informatics* 84.5 (2015), pp. 299–307 (cited on pages 21, 68, 73).

[60] R. W. Grout et al. 'A six-year repeated evaluation of computerized clinical decision support system user acceptability'. In: *International journal of medical informatics* 112 (2018), pp. 74–81 (cited on pages 21, 68).

[61] G. Greenfield, K. Foley, and A. Majeed. 'Rethinking primary care's gatekeeper role'. In: *Bmj* 354 (2016), p. i4803 (cited on page 21).

[62] J. J. G. Wammes et al. 'Is the role as gatekeeper still feasible? A survey among Dutch general practitioners'. In: *Family practice* 31.5 (2014), pp. 538–544 (cited on page 21).

[63] R. Balogh et al. 'Organising healthcare services for persons with an intellectual disability'. In: *Cochrane Database of Systematic Reviews* 4 (2016) (cited on pages 21, 108, 129).

[64] J. Wilkinson et al. '"Sometimes I feel overwhelmed": educational needs of family physicians caring for people with intellectual disability'. In: *Intellectual and developmental disabilities* 50.3 (2012), pp. 243–250 (cited on pages 21, 23).

[65] J. R. Thompson et al. 'Conceptualizing supports and the support needs of people with intellectual disability'. In: *Intellectual and developmental disabilities* 47.2 (2009), pp. 135–146 (cited on page 21).

[66] K. J. Whitt et al. 'Nursing student experiences regarding safe use of electronic health records: a pilot study of the Safety and Assurance Factors for EHR Resilience guides'. In: *CIN: Computers, Informatics, Nursing* 35.1 (2017), pp. 45–53 (cited on pages 21, 68).

[67] J. Adler-Milstein, J. Everson, and S. D. Lee. 'Sequencing of EHR adoption among US hospitals and the impact of meaningful use'. In: *Journal of the American Medical Informatics Association* 21.6 (2014), pp. 984–991 (cited on page 21).

[68] N. Makam A et al. 'Use and satisfaction with key functions of a common commercial electronic health record: a survey of primary care providers'. In: *BMC medical informatics and decision making* 13.1 (2013), p. 86 (cited on page 21).

[69] E. Kihuba et al. 'Assessing the ability of health information systems in hospitals to support evidence-informed decisions in Kenya'. In: *Global health action* 7.1 (2014), p. 24859 (cited on pages 21, 68).

[70] A. Tubaishat and O. M. AL-Rawajfah. 'The use of electronic medical records in jordanian hospitals: a nationwide survey'. In: *CIN: Computers, Informatics, Nursing* 35.10 (2017), pp. 538–545 (cited on page 21).

[71] M. Karimi et al. 'Features of Intellectual Disability Information System: A Comparative Study'. In: *Clin Rev Case Rep* 18 (2018), p. 20 (cited on page 21).

[72] P. S. Sockolow et al. 'Advice for decision makers based on an electronic health record evaluation at a program for all-inclusive care for elders site'. In: *Applied clinical informatics* 2.01 (2011), pp. 18–38 (cited on pages 22, 73).

[73] T. Wang and S. Biedermann. 'Adoption and utilization of electronic health record systems by long-term care facilities in Texas'. In: *Perspectives in Health Information Management/AHIMA, American Health Information Management Association* 9.Spring (2012) (cited on page 22).

[74] B. Cherry et al. 'Factors affecting electronic health record adoption in long-term care facilities'. In: *Journal for Healthcare Quality* 30.2 (2008), pp. 37–47 (cited on page 22).

[75] CCHIT Certification Commission for Healthcare Information Technology. *CCHIT certification for ambulatory electronic health records*. 2020. URL: `https://www.cchit.org/` (cited on page 22).

[76] L. Raymond et al. 'Improving performance in medical practices through the extended use of electronic medical record systems: a survey of Canadian family physicians'. In: *BMC medical informatics and decision making* 15.1 (2015), p. 27 (cited on page 22).

[77] Dutch Ministry of Health Welfare and Sport. *"Act amending the Act on the use of the citizen service number in health care, etc. (client rights in the case of electronic data processing)" [Wijzigingswet Wet gebruik burgerservicenummer in de zorg, enz. (cliëntenrechten bij elektronische verwerking van gegevens)]*. 2016 (cited on page 22).

[78] G. Muller. 'A reference architecture primer'. In: *Eindhoven Univ. of Techn., Eindhoven, White paper* (2012) (cited on pages 23, 68).

[79] C. L. Anderson and R. Agarwal. 'The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information'. In: *Information Systems Research* 22.3 (2011), pp. 469–490 (cited on pages 30, 92).

[80] R. S. Evans. 'Electronic health records: then, now, and in the future'. In: *Yearbook of medical informatics* 25.S 01 (2016), S48–S61 (cited on page 30).

[81] C. C. Agbo, Q. H. Mahmoud, and J. M. Eklund. 'Blockchain technology in healthcare: a systematic review'. In: *Healthcare*. Vol. 7. 2. Multidisciplinary Digital Publishing Institute. 2019, p. 56 (cited on page 30).

[82] L.S. Davis et al. 'Computer-stored medical record'. In: *Computers and Biomedical Research* 1.5 (1968), pp. 452–469 (cited on page 30).

[83] F. Borst et al. 'Happy birthday DIOGENE: a hospital information system born 20 years ago'. In: *International Journal of Medical Informatics* 54.3 (1999), pp. 157–167 (cited on page 30).

[84] B. M. C. Silva et al. 'Mobile-health: A review of current state in 2015'. In: *Journal of biomedical informatics* 56 (2015), pp. 265–272 (cited on page 30).

[85] S. Horng et al. 'Prospective pilot study of a tablet computer in an Emergency Department'. In: *International journal of medical informatics* 81.5 (2012), pp. 314–319 (cited on page 30).

[86] C S. K. Cheung et al. 'Factors associated with adoption of the electronic health record system among primary care physicians'. In: *JMIR medical informatics* 1.1 (2013), e1 (cited on page 30).

[87] A. Wright et al. 'Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems'. In: *Journal of the American Medical Informatics Association* 18.3 (2011), pp. 232–242 (cited on page 30).

[88] W.P. Goh et al. 'Decision support systems for adoption in dental clinics: a survey'. In: *Knowledge-Based Systems* 104 (2016), pp. 195–206 (cited on pages 30, 68).

[89] F. Rezaeibagha, K. T. Win, and W. Susilo. 'A systematic literature review on security and privacy of electronic health record systems: technical perspectives'. In: *Health Information Management Journal* 44.3 (2015), pp. 23–38 (cited on page 30).

[90] F. A. Rahim, Z. Ismail, and G. N. Samy. 'Information privacy concerns in electronic healthcare records: A systematic literature review'. In: *2013 International Conference on Research and Innovation in Information Systems (ICRIIS)*. IEEE. 2013, pp. 504–509 (cited on page 30).

[91] G. da Silva Serapião Leal, W. Guédria, and H. Panetto. 'Interoperability assessment: A systematic literature review'. In: *Computers in Industry* 106 (2019), pp. 111–132 (cited on page 30).

[92] Y. Sun et al. 'Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators'. In: *Journal of medical Internet research* 21.5 (2019), e12522 (cited on page 30).

[93] B. Chaudhry et al. 'Systematic review: impact of health information technology on quality, efficiency, and costs of medical care'. In: *Annals of internal medicine* 144.10 (2006), pp. 742–752 (cited on page 30).

[94] B. Kitchenham et al. 'Systematic literature reviews in software engineering - A systematic literature review'. In: *Information and Software Technology* 51.1 (2009), pp. 7–15 (cited on pages 31, 32, 34, 45).

[95] Cochrane Collaboration. *Cochrane Reviewers' Handbook. Version 4.2.1.* John Wiley and Sons, 2003 (cited on page 32).

[96] J. Tummers, A. Kassahun, and B. Tekinerdogan. ‘Obstacles and features of Farm Management Information Systems: A systematic literature review’. In: *Computers and Electronics in Agriculture* 157 (2019), pp. 189–204 (cited on page 34).

[97] R. van Dinter, B. Tekinerdogan, and C. Catal. ‘Automation of systematic literature reviews: A systematic literature review’. In: *Information and Software Technology* (2021), p. 106589 (cited on page 34).

[98] B. Uzun and B. Tekinerdogan. ‘Model-driven architecture based testing: A systematic literature review’. In: *Information and software technology* 102 (2018), pp. 30–48 (cited on page 34).

[99] C. Avci, B. Tekinerdogan, and I. N. Athanasiadis. ‘Software architectures for big data: a systematic literature review’. In: *Big Data Analytics* 5.1 (2020), pp. 1–53 (cited on pages 34, 102).

[100] D. Kirk, C. Catal, and B. Tekinerdogan. ‘Precision nutrition: A systematic literature review’. In: *Computers in Biology and Medicine* (2021), p. 104365 (cited on page 34).

[101] D. Moher et al. ‘Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement’. In: *Annals of internal medicine* 151.4 (2009), pp. 264–269 (cited on page 35).

[102] A. L. Mendelow. ‘Environmental Scanning–The Impact of the Stakeholder Concept’. In: (1981) (cited on page 45).

[103] M. H. Mobasheri et al. ‘The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study’. In: *BMJ Innovations* 1.4 (2015), pp. 174–181 (cited on page 45).

[104] M. M. Tripathi and N. K Joshi. ‘Big data issues in medical healthcare’. In: *Intelligent Communication, Control and Devices*. Springer, 2018, pp. 1757–1765 (cited on page 45).

[105] M. Zahabi, D. B. Kaber, and M. Swangnetr. ‘Usability and safety in electronic medical records interface design: a review of recent literature and guideline formulation’. In: *Human factors* 57.5 (2015), pp. 805–834 (cited on page 45).

[106] M. Montazeri, R. Khajouei, and M. Montazeri. ‘Evaluating hospital information system according to ISO 9241 part 12’. In: *Digital Health* 6 (2020), p. 2055207620979466 (cited on page 45).

[107] M. Hochgesang et al. ‘Scaling-up health information systems to improve HIV treatment: an assessment of initial patient monitoring systems in Mozambique’. In: *International journal of medical informatics* 97 (2017), pp. 322–330 (cited on page 45).

[108] V. L. Younge, E. M Borycki, and A. W. Kushniruk. ‘On-the-job training of health professionals for electronic health record and electronic medical record use: A scoping review’. In: *Knowledge Management & E-Learning: An International Journal* 7.3 (2015), pp. 436–469 (cited on page 45).

[109] D. Bender and K. Sartipi. ‘HL7 FHIR: An Agile and RESTful approach to healthcare information exchange’. In: *Proceedings of the 26th IEEE international symposium on computer-based medical systems*. IEEE. 2013, pp. 326–331 (cited on pages 46, 86).

[110] World Organization of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians. *ICPC-2: International classification of primary care*. 1998 (cited on pages 46, 86).

[111] World Health Organization. *International classification of diseases for mortality and morbidity statistics (11th Revision)*. 2018. URL: `https://icd.who.int/browse11/l-m/en` (cited on pages 46, 86).

[112] S. Schulz, R. Stegwee, and C. Chronaki. 'Standards in healthcare data'. In: *Fundamentals of Clinical Data Science*. Springer, 2019, pp. 19–36 (cited on page 46).

[113] T. K. L. Schleyer et al. 'Preliminary evaluation of the Chest Pain Dashboard, a FHIR-based approach for integrating health information exchange information directly into the clinical workflow'. In: *AMIA Summits on Translational Science Proceedings* 2019 (2019), p. 656 (cited on page 46).

[114] S. M. H. Dehaghani and N. Hajrahimi. 'Which factors affect software projects maintenance cost more?' In: *Acta Informatica Medica* 21.1 (2013), p. 63 (cited on page 46).

[115] J. Sligo et al. 'A literature review for large-scale health information system project planning, implementation and evaluation'. In: *International journal of medical informatics* 97 (2017), pp. 86–97 (cited on pages 46, 47).

[116] B. Rahimi and V. Vimarlund. 'Methods to evaluate health information systems in healthcare settings: a literature review'. In: *Journal of medical systems* 31.5 (2007), pp. 397–432 (cited on pages 46, 47).

[117] C. Curtain and G. M. Peterson. 'Review of computerized clinical decision support in community pharmacy'. In: *Journal of clinical pharmacy and therapeutics* 39.4 (2014), pp. 343–348 (cited on page 47).

[118] M. J. Ball. 'Hospital information systems: perspectives on problems and prospects, 1979 and 2002'. In: *International journal of medical informatics* 69.2-3 (2003), pp. 83–89 (cited on page 47).

[119] R. S. Janett and P. P. Yeracaris. 'Electronic Medical Records in the American Health System: challenges and lessons learned'. In: *Ciencia & saude coletiva* 25 (2020), pp. 1293–1304 (cited on page 47).

[120] A. J. McLeod Jr and J. G. Clark. 'Using stakeholder analysis to identify users in healthcare information systems research: Who is the real user?' In: *International Journal of Healthcare Information Systems and Informatics (IJHISI)* 4.3 (2009), pp. 1–15 (cited on page 47).

[121] M. Kumar et al. 'Research gaps in routine health information system design barriers to data quality and use in low-and middle-income countries: A literature review'. In: *The International journal of health planning and management* 33.1 (2018), e1–e9 (cited on page 47).

[122] S. S. Kruse et al. 'The use of electronic health records to support population health: a systematic review of the literature'. In: *Journal of medical systems* 42.11 (2018), pp. 1–16 (cited on page 47).

[123] N. Rosenkötter et al. 'Key features of an EU health information system: a concept mapping study'. In: *The European Journal of Public Health* 26.1 (2016), pp. 65–70 (cited on page 47).

[124] World Health Organization. 'Toolkit on monitoring health systems strengthening'. In: *WHO. 2008b. WHO* (2009), pp. 17–3 (cited on page 47).

[125] M. N. Ngafeeson. 'Healthcare information systems opportunities and challenges'. In: *Encyclopedia of Information Science and Technology, Third Edition*. IGI Global, 2015, pp. 3387–3395 (cited on page 47).

[126] A. Roehrs et al. 'Personal health records: a systematic literature review'. In: *Journal of medical Internet research* 19.1 (2017), e13 (cited on page 47).

[127] T. Dybå and T. Dingsøyr. 'Strength of evidence in systematic reviews in software engineering'. In: *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*. ACM, 2008, pp. 178–187 (cited on page 48).

[128] U. A. Bhatti et al. 'Recommendation system using feature extraction and pattern recognition in clinical care systems'. In: *Enterprise Information Systems* 13.3 (2019), pp. 329–351 (cited on page 68).

[129] A. S. O'malley et al. 'Electronic health records and support for primary care teamwork'. In: *Journal of the American Medical Informatics Association* 22.2 (2015), pp. 426–434 (cited on page 68).

[130] A. Friedman et al. 'A typology of electronic health record workarounds in small-to-medium size primary care practices'. In: *Journal of the American Medical Informatics Association* 21.e1 (2014), e78–e83 (cited on page 68).

[131] R. Khajouei, S. Hajesmaeel Gohari, and M. Mirzaee. 'Comparison of two heuristic evaluation methods for evaluating the usability of health information systems'. In: *Journal of biomedical informatics* 80 (2018), pp. 37–42 (cited on page 68).

[132] Y. Park and D. Han. 'Current status of electronic medical record systems in hospitals and clinics in Korea'. In: *Healthcare informatics research* 23.3 (2017), pp. 189–198 (cited on page 68).

[133] G. R. L. Paré et al. 'Electronic health record usage behaviors in primary care medical practices: a survey of family physicians in Canada'. In: *International journal of medical informatics* 84.10 (2015), pp. 857–867 (cited on pages 68, 73).

[134] D.M. Vossebeld et al. 'Development process of a mobile electronic medical record for nurses: a single case study'. In: *BMC medical informatics and decision making* 19.1 (2019), p. 11 (cited on page 68).

[135] D. T. Dunsmuir et al. 'Development of mHealth applications for pre-eclampsia triage'. In: *IEEE journal of biomedical and health informatics* 18.6 (2014), pp. 1857–1864 (cited on page 68).

[136] P. Locatelli et al. 'Health care information systems: architectural models and governance'. In: *Innovative Information Systems Modelling and Techniques* (2012), pp. 71–96 (cited on pages 68, 76).

[137] B. Tekinerdogan, U. Zdun, and A. Babar. *Software Architecture: 10th European Conference, ECSA 2016, Copenhagen, Denmark, November 28–December 2*. Vol. 9839. Springer International Publishing, 2014, Vol. 9839 (cited on page 68).

[138] ISO/IEC/IEEE 42010. *Systems and software engineering–architecture description*. Tech. rep. ISO/IEC/IEEE 42010, 2011 (cited on page 68).

[139] A. Kassahun. 'Aligning business processes and IT of multiple collaborating organisations'. PhD thesis. Wageningen University, 2017 (cited on pages 68, 71).

[140] US Dept. of Defence / Office of the DoD CIO. *Reference Architecture Description*. Tech. rep. June. US Dept. of Defence / Office of the DoD CIO, 2010, p. 22 (cited on page 68).

[141] B. Frandji et al. 'The RICHE reference architecture'. In: *Medical Informatics* 19.1 (1994), pp. 1–11 (cited on pages 68, 86).

[142] F. Wartena et al. 'Continua: The reference architecture of a personal telehealth ecosystem'. In: *The 12th IEEE International Conference on e-Health Networking, Applications and Services*. IEEE. 2010, pp. 1–6 (cited on pages 68, 86).

[143] Extreme networks. *Healthcare Reference Architecture*. Tech. rep. Extreme Networks, 2010 (cited on pages 68, 86).

[144] K. Puscas. *Nationwide Health Information Network Exchange. Enterprise Architecture Overview*. Tech. rep. Nationwide Health Information Network, 2009 (cited on pages 68, 86).

[145] N. Brandara. *Connected Health Reference Architecture*. Tech. rep. WSO2, 2015 (cited on pages 68, 86).

[146] O. M. Kipf. *Reference Architecture for Healthcare – Design concepts*. 2020 (cited on pages 68, 86).

[147] Nictiz. *Domain Reference Model for Hospitals version 2*. Tech. rep. Nationaal ICT Instituut in de Zorg, 2012 (cited on pages 68, 86).

[148] ATOS. *IT Reference Architecture for Healthcare*. Tech. rep. ATOS, 2011 (cited on pages 68, 86).

[149] A. Winter et al. '3LGM2-Modeling to support management of health information systems'. In: *international journal of medical informatics* 76.2-3 (2007), pp. 145–150 (cited on pages 68, 82, 86).

[150] O. Koksal and B. Tekinerdogan. 'Obstacles in Data Distribution Service Middleware: A Systematic Review'. In: *Future Generation Computer Systems* 68 (2017), pp. 191–210 (cited on page 70).

[151] M. van Geest, B. Tekinerdogan, and C. Catal. 'Design of a reference architecture for developing smart warehouses in industry 4.0'. In: *Computers in Industry* 124 (2020), p. 103343 (cited on pages 70, 72).

[152] B. Tekinerdogan and M. Aksit. 'Classifying and evaluating architecture design methods'. In: *Software Architectures and Component Technology*. Springer, 2002, pp. 3–27 (cited on page 70).

[153] P. Clements et al. *Documenting software architectures: views and beyond (2nd Edition)*. Addison-Wesley Professional; 2 edition (October 15, 2010), 2010 (cited on pages 70, 76, 80, 81, 86).

[154] J. Tummers, A. Kassahun, and B. Tekinerdogan. 'Reference architecture design for farm management information systems: a multi-case study approach'. In: *Precision Agriculture* (2020) (cited on pages 71, 72).

[155] D. Yan-e. 'Research about based-SOA agriculture management information system'. In: *Information and Automation (ICIA), 2012 International Conference on*. IEEE. 2012, pp. 78–82 (cited on page 71).

[156] Y. Murakami et al. 'IFarm: development of cloud-based system of cultivation management for precision agriculture'. In: *Consumer Electronics (GCCE), 2013 IEEE 2nd Global Conference on*. IEEE. 2013, pp. 233–234 (cited on page 71).

[157] A. Kassahun, R. J. M. Hartog, and B. Tekinerdogan. 'Realizing chain-wide transparency in meat supply chains based on global standards and a reference architecture'. In: *Computers and Electronics in Agriculture* 123 (Apr. 2016), pp. 275–291 (cited on page 72).

[158] B. Kitchenham and S. Charters. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Tech. rep. Computer Science Department, Keele University and Department of Computer Science, University of Duram, 2007 (cited on page 73).

[159] R. Riahi et al. 'The value of electronic medical record implementation in mental health care: a case study'. In: *JMIR Medical informatics* 5.1 (2017), e1 (cited on page 73).

[160] J. Lei et al. ‘A comparison of electronic health records at two major Peking University Hospitals in China to United States meaningful use objectives’. In: *BMC medical informatics and decision making* 13.1 (2013), p. 96 (cited on page 73).

[161] W. M. Tierney et al. ‘Assessing the impact of a primary care electronic medical record system in three kenyan rural health centers’. In: *Journal of the American Medical Informatics Association* 23.3 (2016), pp. 544–552 (cited on page 73).

[162] J. Guo et al. ‘Usability evaluation of an electronic medication administration record (eMAR) application’. In: *Applied clinical informatics* 2.02 (2011), pp. 202–224 (cited on page 73).

[163] P. Rajkovic, D. Jankovic, and A. Milenkovic. ‘Developing and deploying medical information systems for Serbian public healthcare: Challenges, lessons learned and guidelines.’ In: *Computer Science and Information Systems.* 10.3 (2013), pp. 1429–1454 (cited on page 73).

[164] C. Tsavatewa, P. F. Musa, and I. Ramsingh. ‘Integration of footprints information systems in palliative care: the case of Medical Center of Central Georgia’. In: *Journal of medical systems* 36.3 (2012), pp. 1511–1521 (cited on page 73).

[165] J. D. Duke et al. ‘Regenstrief Institute’s medical gopher: A next-generation homegrown electronic medical record system’. In: *International journal of medical informatics* 83.3 (2014), pp. 170–179 (cited on page 73).

[166] W. Martinez et al. ‘A Patient-Facing Diabetes Dashboard Embedded in a Patient Web Portal: Design Sprint and Usability Testing’. In: *JMIR human factors* 5.3 (2018), e26 (cited on page 73).

[167] B. Tekinerdogan and K. Öztürk. ‘Feature-driven design of SaaS architectures’. In: *Software Engineering Frameworks for the Cloud Computing Paradigm*. Springer, 2013, pp. 189–212 (cited on page 74).

[168] K.C. Kang et al. *Feature-Oriented Domain Analysis (FODA) Feasibility Study*. Tech. rep. Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst, 1990 (cited on page 74).

[169] J. W. H. Mak and J. P. Clarkson. ‘Towards the design of resilient large-scale engineering systems’. In: *Procedia CIRP* 60 (2017), pp. 536–541 (cited on page 76).

[170] C. H. Kim et al. ‘The complementary use of IDEF and UML modelling approaches’. In: *Computers in Industry* 50.1 (2003), pp. 35–56 (cited on pages 76, 95).

[171] M. K. Choubey. *IT Infrastructure and Management (For the GBTU and MMTU)*. Pearson Education India, 2011 (cited on pages 76, 95).

[172] H. Van Vliet. *Software engineering: principles and practice*. Vol. 3. Wiley New York, 1993 (cited on pages 77, 95).

[173] S. Sancar Gozukara, B. Tekinerdogan, and C. Catal. ‘Obstacles of on-premise enterprise resource planning systems and solution directions’. In: *Journal of Computer Information Systems* (2020), pp. 1–12 (cited on page 77).

[174] R. K. Chellappa and N. Saraf. ‘Alliances, rivalry, and firm performance in enterprise systems software markets: a social network approach’. In: *Information Systems Research* 21.4 (2010), pp. 849–871 (cited on page 77).

[175] N. Pollock and R. Williams. ‘Moving beyond the single site implementation study: How (and why) we should study the biography of packaged enterprise solutions’. In: *Information Systems Research* 23.1 (2011), pp. 1–22 (cited on page 77).

[176] F. Jahn et al. 'Comparing a Japanese and a German hospital information system'. In: *Methods of information in medicine* 48.06 (2009), pp. 531–539 (cited on pages 82–86).

[177] J. Opara-Martins, R. Sahandi, and F. Tian. 'Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective'. In: *Journal of Cloud Computing* 5.1 (2016), pp. 1–18 (cited on page 85).

[178] ISO/IEC/IEEE. *Life cycle processes –Requirements engineering*. Tech. rep. ISO/IEC/IEEE, 2011, pp. 1–94 (cited on page 86).

[179] Open Group. *ArchiMate 2.0 Specification: Open Group Standard*. Van Haren Publishing, 2012 (cited on page 86).

[180] P. Saha. *Advances in government enterprise architecture*. IGI Global, 2008 (cited on page 86).

[181] IEEE. *IEEE 1471, Standard for describing the architecture of a "software-intensive system"*. Tech. rep. Institute of Electrical and Electronic Engineers, 2000 (cited on page 86).

[182] J. Tummers et al. 'State of the practice of health information systems: a survey study amongst health care professionals in intellectual disability care'. In: *BMC health services research* 21.1 (2021), pp. 1–9 (cited on pages 92, 108).

[183] J. Warren and N. Marz. *Big Data: Principles and best practices of scalable realtime data systems*. Simon and Schuster, 2015 (cited on page 92).

[184] B. Tekinerdogan and A. Oral. 'Performance Isolation in Cloud-Based Big Data Architectures'. In: *Software architecture for big data and the cloud*. Elsevier, 2017, pp. 127–145 (cited on page 92).

[185] P. Clements et al. 'Documenting software architectures: views and beyond'. In: *25th International Conference on Software Engineering, 2003. Proceedings.* IEEE. 2003, pp. 740–741 (cited on pages 92, 95).

[186] J. Tummers et al. 'Designing a reference architecture for health information systems'. In: *BMC Medical Informatics and Decision Making* 21.1 (2021), pp. 1–14 (cited on pages 92, 95, 96, 99, 100).

[187] H. Chalghoumi et al. 'Information privacy for technology users with intellectual and developmental disabilities: why does it matter?' In: *Ethics & Behavior* 29.3 (2019), pp. 201–217 (cited on page 92).

[188] S. H. Spewak and M. Tiemann. 'Updating the enterprise architecture planning model'. In: *Journal of Enterprise Architecture* 2.2 (2006), pp. 11–19 (cited on page 92).

[189] P. Runeson and M. Höst. 'Guidelines for conducting and reporting case study research in software engineering'. In: *Empirical Software Engineering* 14.2 (Apr. 2009), pp. 131–164 (cited on page 94).

[190] A. Uijen, H. Bor, and K. van Boven. 'FaMe-Net: twee oude registratienetwerken in een nieuw jasje'. In: *Tijdschrift voor gezondheidswetenschappen* 93.8 (2015), pp. 286–287 (cited on page 95).

[191] P. Pääkkönen and D. Pakkala. 'Reference architecture and classification of technologies, products and services for big data systems'. In: *Big data research* 2.4 (2015), pp. 166–186 (cited on pages 96, 102, 106).

[192] H. Schers et al. 'The COVID-19 pandemic in Nijmegen, the Netherlands: changes in presented health problems and demand for primary care'. In: *The Annals of Family Medicine* 19.1 (2021), pp. 44–47 (cited on page 97).

[193] J. M. Groeneveld et al. 'Sex differences in incidence of respiratory symptoms and management by general practitioners'. In: *Family Practice* 37.5 (2020), pp. 631–636 (cited on page 97).

[194] S. J. Faber et al. 'Preventing acute kidney injury in high-risk patients by temporarily discontinuing medication–an observational study in general practice'. In: *Bmc nephrology* 20.1 (2019), pp. 1–7 (cited on page 97).

[195] M. D. Wilkinson et al. 'The FAIR Guiding Principles for scientific data management and stewardship'. In: *Scientific data* 3.1 (2016), pp. 1–9 (cited on page 97).

[196] R. Ras, C. Schuengel, and P. J. C. M. Embregts. 'Zorgdata in onderzoek voor mensen met verstandelijke beperkingen'. In: (2022) (cited on page 97).

[197] M. Aminzade. 'Confidentiality, integrity and availability–finding a balanced IT framework'. In: *Network Security* 2018.5 (2018), pp. 9–11 (cited on page 101).

[198] S. Samonas and D. Coss. 'The CIA strikes back: Redefining confidentiality, integrity and availability in security.' In: *Journal of Information System Security* 10.3 (2014) (cited on page 101).

[199] A. Tchernykh et al. 'Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability'. In: *Journal of Computational Science* 36 (2019), p. 100581 (cited on page 101).

[200] C. A. Salma, B. Tekinerdogan, and I. N. Athanasiadis. 'Domain-driven design of big data systems based on a reference architecture'. In: *Software Architecture for Big Data and the Cloud*. Elsevier, 2017, pp. 49–68 (cited on page 102).

[201] V. Menger et al. 'DEDUCE: A pattern matching method for automatic de-identification of Dutch medical text'. In: *Telematics and Informatics* 35.4 (2018), pp. 727–736 (cited on pages 107, 129).

[202] P. Voigt and A. Von dem Bussche. 'The eu general data protection regulation (gdpr)'. In: *A Practical Guide, 1st Ed., Cham: Springer International Publishing* 10.3152676 (2017), pp. 10–5555 (cited on page 107).

[203] Dutch minister of Health Welfare and Sport. 'Wet medisch–wetenschappelijk onderzoek met mensen (wettekst)'. In: (2008) (cited on page 107).

[204] D. A. Ervin et al. 'Healthcare for persons with intellectual and developmental disability in the community'. In: *Frontiers in public health* 2 (2014), p. 83 (cited on pages 108, 129).

[205] M. Rousseau et al. 'Healthcare of Persons With Complex Developmental Disabilities From Three European Experiences: France, Italy, and Norway'. In: *Journal of Policy and Practice in Intellectual Disabilities* 18.4 (2021), pp. 273–278 (cited on pages 108, 129).

[206] World Health Organization. *Novel Coronavirus – China update*. 2020. URL: `https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/` (cited on page 114).

[207] World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. 2020. URL: `https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020` (cited on page 114).

[208] World Health Organization. *Coronavirus disease (COVID-19) Pandemic*. 2020. URL: `https://www.who.int/emergencies/diseases/novel-coronavirus-2019` (cited on page 114).

[209] C. Lai et al. 'Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges'. In: *International journal of antimicrobial agents* (2020), p. 105924 (cited on page 114).

[210] G Glover et al. 'Mortality in people with intellectual disabilities in England'. In: *Journal of Intellectual Disability Research* 61.1 (2017), pp. 62–74 (cited on page 114).

[211] K. McIntosh and S. Perlman. 'Coronaviruses, including severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS)'. In: *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, Updated Edition. 8th ed. Philadelphia, PA: Elsevier Saunders* (2015) (cited on page 114).

[212] B. Aljaber et al. 'Document clustering of scientific texts using citation contexts'. In: *Information Retrieval* 13.2 (2010), pp. 101–131 (cited on page 114).

[213] L. Amador Penichet, D. Magdaleno Guevara, and M. M. Garc\'\ia Lorenzo. 'New Similarity Function for Scientific Articles Clustering based on the Bibliographic References'. In: *Computación y Sistemas* 22.1 (2018), pp. 93–102 (cited on page 114).

[214] CORD-19. *COVID-19 Open Research Dataset (CORD-19)*. 2020. URL: `https://pages.semanticscholar.org/coronavirus-research` (cited on page 116).

[215] White house Office of Science and Technology policy. *Call to Action to the Tech Community on New Machine Readable COVID-19 Dataset*. 2020. URL: `https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/` (cited on page 116).

[216] The United States National Library of Medicine at the National Institutes of Health. *Pubmed*. 2020. URL: `https://ncbi.nl.nih.gov/pubmed/` (cited on page 116).

[217] World Health Organization. *WHO Covid-19 database*. 2020. URL: `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov` (cited on page 116).

[218] Cold Spring Harbor Laboratory. *bioRxiv*. 2020. URL: `bioRxiv.org` (cited on page 116).

[219] Cold Spring Harbor Laboratory. *medRxiv*. 2020. URL: `medRxiv.org` (cited on page 116).

[220] Kaggle. *Kaggle homepage*. 2020. URL: `https://www.kaggle.com` (cited on page 116).

[221] Expertisepunt verstandelijke beperkingen. *Syndromen*. 2020. URL: `https://www.expertisepuntvb.nl/matige-en-ernstige-verstandelijke-beperking/syndromen%0A` (cited on page 116).

[222] M. Hearst. 'What is text mining'. In: *SIMS, UC Berkeley* 5 (2003) (cited on page 116).

[223] Python Software Foundation. *Python Language Reference, version 3.6.6*. 2020. URL: `https://www.python.org/psf/` (cited on page 117).

[224] Pandas community. *Pandas version 1.0.3*. 2020. URL: `https://pandas.pydata.org` (cited on page 117).

[225] scikit-learn community. *Scikit-learn Version 0.22.2*. 2020. URL: `scikit-learn.org` (cited on page 117).

[226] A. Rajaraman and J.D. Ullman. *Mining of massive datasets*. Cambridge University Press, 2011 (cited on page 117).

[227] A. K. Jain. 'Data clustering: 50 years beyond K-means'. In: *Pattern recognition letters* 31.8 (2010), pp. 651–666 (cited on page 117).

[228] T. M. Kodinariya and P. R. Makwana. 'Review on determining number of Cluster in K-Means Clustering'. In: *International Journal* 1.6 (2013), pp. 90–95 (cited on page 118).

[229] A. Ng. 'Clustering with the k-means algorithm'. In: *Machine Learning* (2012) (cited on page 118).

[230] S. Wold, K. Esbensen, and P. Geladi. 'Principal component analysis'. In: *Chemometrics and intelligent laboratory systems* 2.1-3 (1987), pp. 37–52 (cited on page 118).

[231] Allen Institute for AI. *CORD-19 Explorer*. 2020. URL: `https://cord-19.apps.allenai.org` (cited on page 122).

[232] D. R. Swanson. 'Literature-based discovery? The very idea'. In: *Literature-based discovery*. Springer, 2008, pp. 3–11 (cited on page 122).

[233] Y. Sebastian, E. Siew, and S. O. Orimaye. 'Emerging approaches in literature-based discovery: techniques and performance review'. In: *The Knowledge Engineering Review* 32 (2017) (cited on page 122).

[234] K. Fujita. 'Finding linkage between sustainability science and technologies based on citation network analysis'. In: *2012 Fifth IEEE International Conference on Service-Oriented Computing and Applications (SOCA)*. IEEE. 2012, pp. 1–6 (cited on page 122).

[235] R. N. Kostoff. 'Literature-related discovery: common factors for Parkinson's Disease and Crohn's Disease'. In: *Scientometrics* 100.3 (2014), pp. 623–657 (cited on page 122).

[236] A. Bahga and V. K. Madisetti. 'A cloud-based approach for interoperable electronic health records (EHRs)'. In: *IEEE Journal of Biomedical and Health Informatics* 17.5 (2013), pp. 894–906 (cited on page 127).

[237] M. Jayaratne et al. 'A data integration platform for patient-centered e-healthcare and clinical decision support'. In: *Future Generation Computer Systems* 92 (2019), pp. 996–1008 (cited on pages 127, 129).

[238] T. Iacono and V. Murray. 'Issues of informed consent in conducting medical research involving people with intellectual disability'. In: *Journal of Applied Research in Intellectual Disabilities* 16.1 (2003), pp. 41–51 (cited on page 128).

[239] P. Ho et al. 'Addressing challenges in gaining informed consent for a research study investigating falls in people with intellectual disability'. In: *British Journal of Learning Disabilities* 46.2 (2018), pp. 92–100 (cited on page 128).

[240] M. Cuypers et al. 'Disparities in cancer-related healthcare among people with intellectual disabilities: A population-based cohort study with health insurance claims data'. In: *Cancer medicine* 9.18 (2020), pp. 6888–6895 (cited on page 128).

[241] M. Cuypers et al. 'Mortality of people with intellectual disabilities during the 2017/2018 influenza epidemic in the Netherlands: Potential implications for the COVID-19 pandemic'. In: *Journal of Intellectual Disability Research* 64.7 (2020), pp. 482–488 (cited on page 128).

[242] V. S. Anoop. 'Text Mining and Natural Language Processing for Health Informatics: Recent Trends and the Way Forward'. In: *Machine Learning and Data Analytics for Predicting, Managing, and Monitoring Disease* (2021), pp. 111–121 (cited on page 128).

[243] G. Wijnholds and M. Moortgat. 'SICKNL: A dataset for Dutch natural language inference'. In: *arXiv preprint arXiv:2101.05716* (2021) (cited on page 128).

[244] J. M. Nobel et al. 'Natural language processing in Dutch free text radiology reports: challenges in a small language area staging pulmonary oncology'. In: *Journal of digital imaging* 33.4 (2020), pp. 1002–1008 (cited on page 128).

[245] *Long-term care workforce: Employment and working conditions*. Tech. rep. (cited on page 129).

[246] P. Deshpande et al. 'Big data integration case study for radiology data sources'. In: *2018 IEEE Life Sciences Conference (LSC)*. IEEE. 2018, pp. 195–198 (cited on page 129).

[247] A. R. H. Fischer, H. Tobi, and A. Ronteltap. 'When natural met social: a review of collaboration between the natural and social sciences'. In: *Interdisciplinary Science Reviews* 36.4 (2011), pp. 341–358 (cited on page 129).

[248] R. K. E. Bellamy et al. 'AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias'. In: *IBM Journal of Research and Development* 63.4/5 (2019), pp. 4–1 (cited on page 130).

[249] E. Kazim and A. S. Koshiyama. 'A high-level overview of AI ethics'. In: *Patterns* 2.9 (2021), p. 100314 (cited on page 130).

[250] A. Jobin, M. Ienca, and E. Vayena. 'The global landscape of AI ethics guidelines'. In: *Nature Machine Intelligence* 1.9 (2019), pp. 389–399 (cited on page 130).

[251] H. Sanders and J. Saxe. 'Garbage in, garbage out: how purportedly great ML models can be screwed up by bad data'. In: *Proceedings of Blackhat* 2017 (2017) (cited on page 130).

[252] M. Dzieżyc et al. 'Can we ditch feature engineering? end-to-end deep learning for affect recognition from physiological sensor data'. In: *Sensors* 20.22 (2020), p. 6535 (cited on page 130).

[253] N. C. Thompson, S. Ge, and G. F. Manso. 'The importance of (exponentially more) computing power'. In: *arXiv preprint arXiv:2206.14007* (2022) (cited on page 130).

[254] A. Shehabi et al. 'United states data center energy usage report'. In: (2016) (cited on page 130).

[255] *Andrea consortium*. URL: `https://www.andrea-cloud.eu/` (cited on page 131).

[256] D. W. Nixon. 'Down syndrome, obesity, alzheimer's disease, and cancer: A brief review and hypothesis'. In: *Brain sciences* 8.4 (2018), p. 53 (cited on page 132).

EPILOGUE

Summary

A wide variety of health information systems are used in the care for individuals with intellectual disabilities. In these health information systems, much is recorded about the individuals. For example, personal care aides register in an electronic client dossier how the client was doing on that day, while physical therapists register treatments on the same day in their electronic patient dossier. Also, a general practitioner or a doctor for the intellectually disabled (ID physician) is visited, whom both use another health information system to register their treatment.

All these routinely collected registrations together create a unique and very complete picture about people with intellectual disabilities. Especially when they receive residential care from a care organization, there is a record for almost every day, for their whole life. This routinely collected longitudinal data, therefore, has enormous potential to be re-used for research and to make the care of people with intellectual disabilities better. The objective of this thesis is to propose a framework for a big data system that combines data from different care information systems and facilitates data re-use for analysis.

This thesis consists of seven chapters that work towards the goal stated above. An overview of the chapters is available in Figure 1. After an introduction in Chapter 1, Chapter 2 gives insight into which systems are used in the Dutch care for people with intellectual disabilities, achieved through a survey study with 328 respondents. In total 52 unique systems were identified with a set of 17 most used functionalities and 10 most common problems. Furthermore, caregivers' satisfaction with their systems fluctuated. In chapter 3 the literature was analyzed with a systematic literature review on health information systems. Based on 136 articles, most of them from the hospital domain, 41 stakeholders, 72 functionalities, and 69 obstacles were distilled.

Based on the information from practice (chapter 2) and the state-of-the-art from the literature (chapter 3), a software architecture for health information systems was developed in chapter 4. Just like for a house, you can create an architecture for software that specifies where the system interacts with, what modules the system consists of, how these modules interact with each other, and how the system is structured in terms of servers and databases. The reference architecture in Chapter 4 thus shows the similarities between different healthcare information systems using four diagrams.

The ultimate goal was to make routinely collected data in intellectual disability care, spread across different systems, available for research. From the previous chapters, it became clear that the data that is collected is very diverse, and many privacy and security issues need to be resolved before it can be re-used. In order to facilitate the re-use of data in intellectual disability care, chapter 5 provides an architecture for a big data system. In this chapter, privacy and security issues are mentioned, and solutions are proposed in the form of an architecture for a big data analysis system.

Much of the data collected in intellectual disability care is text. As a brief example of how textual data can be used for analysis, and how large data files can be analyzed, an example is provided in chapter 6 using the CORD-19 data set. This data set contains tens of thousands of articles on COVID-19. Using text mining techniques, there was searched for articles about people with intellectual disabilities and a clustering algorithm was used to group them into five clusters that researchers can use for further research.

Based on the above-mentioned chapters, it was learned that the data landscape in the care of people with intellectual disabilities is complicated, and multiple conditions need to be met before it can be analyzed. However, the enormous potential for the re-use of this data keeps coming up. Future research involves building a big data system following the architecture as presented in chapter 5 such that analyses can be done on this dataset with enormous potential.

Based on the knowledge gained about the ID care sector, the HISs used, and the problems involved in re-using data, POS^T4V is proposed for a big data system in ID care. This abbreviation stands for: **P**rivacy, **O**rganisation of care, **S**ecurity, enabled by **T**echnical solutions to be able to act with **V**olume, **V**elocity, **V**ariety, **V**eracity of data in ID care.

When all stakeholders around the care of people with intellectual disabilities put their proverbial shoulder to the wheel on this, from the people on the work floor, people with intellectual disabilities, software vendors, governments, to the professional associations, I am convinced that after this research we can make the care of people with intellectual disabilities better with data.

Joep Tummers, 2022

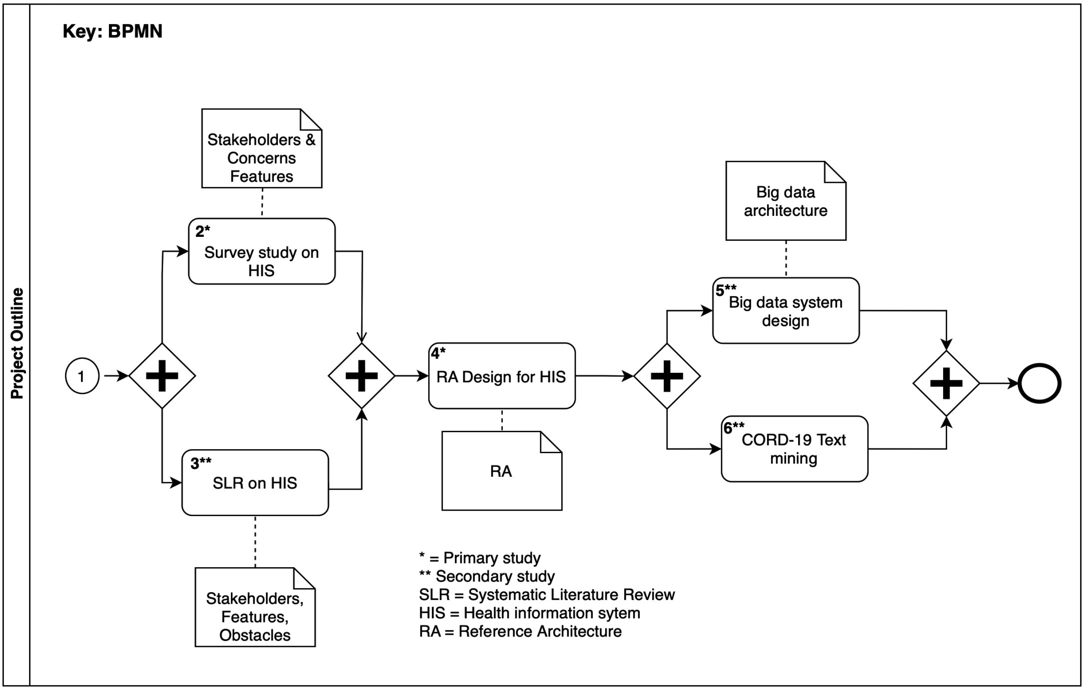


Figure 1: Thesis outline. Numbers correspond to chapter numbers. The start and end events represent the introduction and the discussion, respectively.

Samenvatting

In de zorg voor mensen met een verstandelijke beperking wordt een grote verscheidenheid aan zorg informatie systemen gebruikt. In deze zorg informatie systemen wordt veel geregistreerd over individuen met een verstandelijke beperking. Persoonlijk begeleiders registreren bijvoorbeeld in een elektronisch cliënten dossier hoe het met de cliënt is gegaan op die dag, terwijl fysiotherapeuten op dezelfde dag behandelingen bij deze persoon registreren in hun elektronisch patiënten dossier. Ook komt deze persoon bij een huisarts of een arts voor verstandelijk gehandicapten (AVG-arts) welke een ander zorg informatie systeem gebruikt om hun behandeling te registeren.

Al deze routinematig verzamelde registraties samen zorgen voor een uniek en zeer compleet beeld over mensen met een verstandelijke beperking. Zeker wanneer ze bij een zorgorganisatie woonzorg krijgen, is er bijna voor elke dag een registratie, voor heel hun leven. Deze routinematig verzamelde longitudinale data hebben hierom een enorm potentieel om hergebruikt te worden voor onderzoek en om de zorg voor mensen met een verstandelijke beperking beter te maken. Het doel van deze thesis is om een raamwerk voor te stellen voor een big data systeem welke data combineert uit verschillende zorg informatie systemen en data hergebruik voor analyses faciliteert.

Deze thesis bestaat uit zeven verschillende hoofdstukken welke samen toewerken richting het hierboven gestelde doel. Een overzicht van de hoofdstukken is beschikbaar in Figuur 1. Na een introductie in Hoofdstuk 1 geeft hoofdstuk 2 inzicht in welke systemen gebruikt worden in de Nederlandse zorg voor mensen met een verstandelijke beperking door middel van een enquête studie met 328 respondenten. In totaal kwamen hier 52 unieke systemen uit met een set van 17 meest gebruikte functionaliteiten en 10 meest voorkomende problemen. Verder was de tevredenheid van de zorgverleners over hun systeem fluctuerend. In hoofdstuk 3 is er naar de literatuur gekeken met een systematisch literatuur onderzoek naar zorg informatie systemen. Aan de hand van 136 artikelen, waarvan de meeste uit het ziekenhuis domein, zijn hier 41 belanghebbenden, 72 functionaliteiten en 69 obstakels uit gedestilleerd. Gebaseerd op de informatie uit de praktijk (hoofdstuk 2) en de state-of-the-art uit de literatuur (hoofdstuk 3) is er een software architectuur voor zorg informatie systemen ontwikkeld in hoofdstuk 4. Net als voor een huis, kun je voor software een architectuur maken welke aangeeft waarmee het systeem interacties heeft, uit welke modules het systeem bestaat, hoe deze modules met elkaar interacteren, en hoe het systeem gestructureerd is qua servers en databases. De referentie architectuur in hoofdstuk 4 geeft hiermee de overeenkomsten tussen de verschillende zorg informatie systemen weer aan de hand van vier diagrammen. Het uiteindelijke doel was om routinematige verzamelde data in de verstandelijk beperkten zorg, verspreid over verschillende systemen, beschikbaar te maken voor onderzoek. Uit de voorgaande hoofdstukken kwam naar voren dat de data welke verzameld wordt zeer divers is en er veel privacy en veiligheid problemen zijn welke opgelost moeten worden voordat deze hergebruikt kan gaan worden. Voor het faciliteren van hergebruik van data in de verstandelijk beperkten zorg geeft hoofdstuk 5 een architectuur voor een big data systeem. In dit hoofdstuk worden privacy en security vraagstukken benoemd en worden oplossingen aangedragen in de vorm van een architectuur voor een big data analyse systeem.

Veel van de data verzameld in de verstandelijk beperkten zorg is tekst. Als kort voorbeeld van hoe tekstuele data ingezet kan worden voor analyses, en hoe grote data bestanden geanalyseerd kunnen worden, geeft hoofdstuk 6 een voorbeeld aan de hand van de CORD-19 data set. Deze dataset bevat tienduizenden artikelen over COVID-19. Met behulp van tekst mining technieken is gezocht welke artikelen over mensen met een verstandelijke beperking gaan, en zijn deze met een cluster algoritme gegroepeerd in vijf clusters welke onderzoekers kunnen gebruiken voor verder onderzoek.

Aan de hand van de hierboven genoemde hoofdstukken is ondervonden dat het data landschap in de zorg voor mensen met een verstandelijke beperking ingewikkeld is en er meerdere voorwaarden zijn voordat deze geanalyseerd kan worden. Echter komt steeds weer de enorme potentie voor het hergebruik van deze data naar boven. Toekomstig onderzoek behelst het bouwen van een big data systeem volgende de architectuur zoals benoemd in hoofdstuk 5 zodat analyses gedaan kunnen worden op deze dataset met enorme potentie. Op basis van de opgedane kennis over de ID-zorgsector, de gebruikte zorg informatie systemen en de problemen bij hergebruik van gegevens wordt POS^T4V voorgesteld voor een big data systeem in de ID-zorg. Deze afkorting staat voor: **P**rivacy, **O**rganisatie van de zorg, veiligheid (**S**ecurity), mogelijk gemaakt door **T**echnische oplossingen om te kunnen gaan met de 4Vs van big data in de ID-zorg (Volume, Variety, Veracity, Velocity)

Wanneer alle belanghebbenden rondom de zorg voor mensen met een verstandelijke beperking hier hun spreekwoordelijke schouders onder zetten, van de mensen op de werkvloer, mensen met een verstandelijke beperking, software leveranciers, overheden, tot aan de beroepsverenigingen, ben ik er na dit onderzoek van overtuigd dat we de zorg voor mensen met een verstandelijke beperking beter kunnen maken met behulp van data.

Joep Tummers, 2022

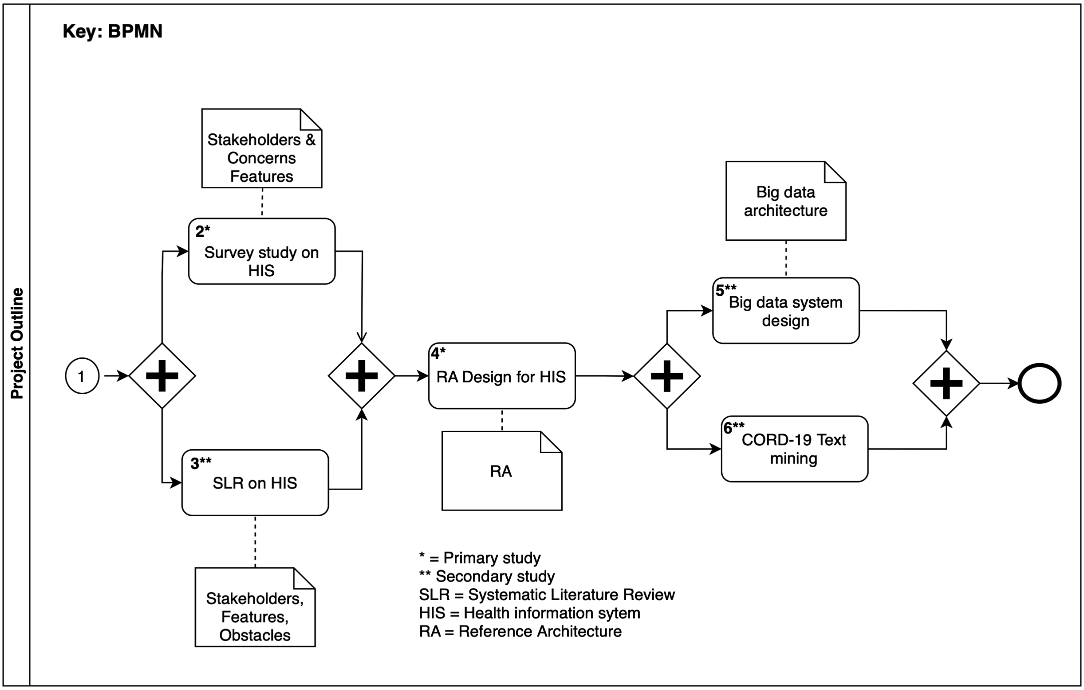


Figure 1: Overzicht van thesis. Nummers corresponderen met hoofdstukken in de thesis.

About the Author

Joep Tummers was born on the 6th of July 1995 as the oldest of two sons in the village of Limmen, in Noord-Holland. As a young child Joep was already interested in the agricultural sector and told his parents he wanted to become a farmer. After finishing his high school in Castricum he followed his passion for the agricultural sector and started studying Agricultural Engineering at the University of Wageningen in 2013. After a minor at the University of Illinois in the U.S.A he finished his bachelor in 2016 and started his master Biosystems Engineering.

With a specialization in automotive engineering at the University of Eindhoven and two published journal articles on farm management information systems from this MSc thesis he graduated in 2018. Directly after finishing his MSc he started his PhD on data science for the care for people with an intellectual disability.

Currently Joep is working as a part-time lecturer at the Information Technology group at the Wageningen University and part-time as a digital innovation engineer at the agricultural machine manufacturer J.O.Z. in Westwoud in Noord-Holland. He is back to his roots, living in Limmen together with his girlfriend Nienke and their cat.

Education certificate

Hubertus Johannes Martinus Tummers
Wageningen School of Social Sciences (WASS)
Completed Training and Supervision Plan

Wageningen School
of Social Sciences

Name of the learning activity	Department/Institute	Year	ECTS*
A) Project related competences			
A1 Managing a research project			
WASS Introduction Course	WASS	2019	1
'AI for Intellectual Disability care: Architecting Health information Systems in Intellectual Disability Care'	WASS PhD Day	2021	0.5
Writing research proposal	WASS	2019	4
"More than an Apple a Day"	IASSIDD Conference	2019	1
"Research data on people with ID: How FA(I)R are we?"	IASSIDD Conference	2019	1
"Big data: An inclusive science platform to improve health and healthcare"	IASSIDD Conference	2021	1
A2 Integrating research in the corresponding discipline			
FTE-35306, Machine Learning	WUR	2019	6
Advanced Online & Onsite Course on Data Science & Machine Learning	ICAS	2021	8
Health Information Literacy for Data Analytics Specialization	UC Davis (online)	2021	1.7
Machine Learning	Stanford University (Online)	2018	2.8
Intellectual Disability Healthcare	University of Queensland (Online)	2020	1.1
Web of Data	EIT digital (online)	2019	0.7
B) General research related competences			
B1 Placing research in a broader scientific context			
Philosophy of Social Science	WASS	2020	3
Ethics of artificial intelligence	University of Helsinki (Online)	2021	2
Introduction to Animal Behaviour	WUR (Online)	2022	1
B2 Placing research in a societal context			
The art of presenting science	Radboud University	2021	1.5
C) Career related competences/personal development			
C1 Employing transferable skills in different domains/careers			
Reviewing a scientific paper	WGS	2019	0.1
Mobilising your scientific network	WGS	2019	1
Supervising BSc and MSc Thesis students	WGS	2020	0.6
Teaching and Supervision	INF, WUR	2019	4
Total			**42**

*One credit according to ECTS is on average equivalent to 28 hours of study load

Acknowledgements

Nienke, jij allereerst duizendmaal dank. Je hebt me door deze periode heen gesleurd kun je wel zeggen. Het doen van deze PhD was voor mij niet altijd even makkelijk, maar door jouw steun is het me gelukt. Vader, Moeder, Maarten, dank voor jullie steun gedurende deze tijd, het was fijn om weer bij jullie thuis te komen in het weekend na een week Ph-Den.

Hilde, jou ben ik veel dank verschuldigd. Niet alleen voor je altijd mega goede feedback, maar ook voor het zijn van een luisterend oor. Af en toe een wandeling, of bij jullie thuis eerst wat werken en daarna uitgebreid aan de koffie met Rob erbij heeft mij veel goeds gedaan. Bedir, door jou ben ik in de gelegenheid om dit stukje te schrijven, dank hiervoor! Onze discussies gaven mij altijd weer inspiratie om verder te gaan, maar ook de meer serieuze gesprekken heb ik altijd als erg fijn ervaren. Cagatay, thank you very much for your insights into the beautiful world of data sciences. You were always the first to respond to my (many) requests for feedback with relatively easy to solve comments :-).

Geraline, ik zal je deze keer sparen en niet “roasten”. Dank voor het mij in staat stellen om deze PhD af te ronden, inclusief het presenteren en bezoeken van conferenties. Met jou gedachtes wisselen over de mogelijkheden van data in de zorg voor mensen met een verstandelijke beperking was altijd weer erg leuk en nuttig. Verder was onze trip naar Orlando uiteraard een hoogtepunt eerste klas. Ook onze gesprekken met het ministerie hebben ik altijd zeer leerzaam en leuk gevonden.

My dear INF PhD colleagues Dilli, Christos, Laura, especially during the first year when everybody was still in the office, I really enjoyed working together. Our table tennis matches were of a very high level.

Iemand die wel altijd op kantoor was, was onze Sjon Kik. Sjon, dank voor de gezellige koffie momenten, onze diepgaande discussies over het academische systeem, het zeiken op van alles, en het zuipen van enorme bulten bier. Ook dank aan “Kollekas” Thijs en Joost, het was altijd weer een eer om met jullie van gedachten te wisselen. Also, many Thanks / Bedankt / Grazie mille / Thank you in Chinese to the other ORL PhDs for adopting me during some of your outings.

Dan zijn we bij het Kluppie uit Nijmegen aanbeland. Allereerst dank naar de GMVB PhDs / S.T.A.R. Lads (Milou, Marian, Masha, Kim, Julia, Anne, Kristel, Amina) ! Onze leuke borrels en activiteiten samen waren altijd iets waar ik enorm naar uit kon kijken. Maarten, bedankt voor onze gezellige trips naar het V.K. en Finland, ik heb er enorm van genoten en veel van opgestoken. Hopelijk kunnen we in de toekomst blijven samenwerken. Ook het ressie van GMVB wil ik enorm bedanken voor de gezellige activiteiten en het uitwisselen van ideeën. Special thanks aan onze bowlingborrel koning (Tim) en koningin (Bianca). Bianca, jij uiteraard ook nog enorm bedankt voor je goede input in mijn proefschrift. Jouw kennis kwam altijd enorm goed van pas.

Vrienden van Agrotechnologie (Pieter, Jeroen, Roel, Jari, Arjan) in het eerste jaar nog dank voor de gezelligheid samen in Wageningen. Vooral de etentjes met jou Pieter waren altijd van hoog creatief niveau. Daarnaast bedankt voor de leuke uitjes en borrels gedurende het jaar, maar ook het samen beleven van de jammer genoeg minder leuke momenten in het leven.

Verder nog dank naar mijn vrienden van thuisthuis. Het was altijd erg fijn om met jullie over dingen te brainstormen tijdens een goede borrel. Joris, jij vooral bedankt. Tijdens onze wintersporten en andere (minder) leuke momenten ben je me altijd tot steun geweest, merci!

Verder nog dank aan Wilbert en Karla voor jullie luisterend oor, Collega's van JOZ voor het voorzien van afleidende werkzaamheden, en Heeren XVII voor het organiseren van leuke en leerzame activiteiten.

The research described in this thesis was financially supported by the Dutch Ministry of Health, Welfare and sport.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged

Cover design by DALL·E 2 (OpenAI)
Printed by Proefschrifmaken