OPEN ACCESS

Research article

A process model for acquiring international administrative routine data for health services research

Konzeption eines Prozessmodells zur Akquise von administrativen Routinedaten für die Forschung

Abstract

Objectives: To describe a practical and standardized approach for acquiring international administrative routine data from different data owners for research.
Methods: Best practice approach based on the experiences gained during the EU-funded ADVOCATE ("Added Value for Oral Care") project that involved the collection of routinely collected administrative data from health insurance providers, health funds or health authorities in six European countries.
Results: A general process for data acquisition that contains four phases was developed: First, the conditions for data usage and access are determined. These conditions are subsequently tested by sharing and analyzing a data sample (quality and validity audit). After optimizing the process model, full-scale data access and analysis are performed.
Conclusions: The general data acquisition approach has successfully been applied in the ADVOCATE project to acquire claims data from eight data owners, which prescribed different usage conditions in each case. The approach aims to make a contribution to a standardized process model for acquiring administrative routine data for research and providing researchers with a methodological framework.

Keywords: administrative data, health information technology, health services research, secondary data analysis, best practice

Zusammenfassung

Ziel: Konzeption eines anwendbaren und standardisierten Ansatzes zur Akquise internationaler administrativer Routinedaten von verschiedenen Dateneigentümern für die Forschung.
Methoden: Best-Practice-Ansatz auf Grundlage der Erfahrungen aus dem EU-Projekt ADVOCATE (Added Value for Oral Care), bei dem administrative Routinedaten von Krankenkassen, Krankenversicherungen oder Gesundheitsbehörden aus sechs europäischen Ländern erhoben wurden.
Ergebnisse: Es wurde ein allgemeines, vierstufiges Verfahren zur Datenakquise entwickelt: Zunächst werden die Bedingungen für die Datennutzung und den Datenzugriff festgelegt. Diese Bedingungen werden anschließend durch den Austausch und die Analyse einer Stichprobe mit anschließender Datenqualitätsprüfung getestet. Nach der Optimierung des Prozesses erfolgt der vollständige Datenzugriff und die Analyse.
Schlussfolgerungen: Das entwickelte Verfahren zur Datenakquise wurde erfolgreich im ADVOCATE- Projekt angewandt, um administrative Routinedaten von acht Dateneigentümern zu akquirieren, die jeweils unterschiedliche Nutzungsbedingungen vorschrieben. Ziel des Ansatzes ist es, einen Beitrag zu einem standardisierten Verfahren zur Akquise von Routinedaten für die Forschung zu leisten.

Christian Haux[1]
Kasper Rosing[2]
Petra Knaup[1]
Stefan Listl[3,4]
Olivier Kalmus[4]

1 Institute for Medical Biometry and Informatics, Heidelberg University, Heidelberg, Germany

2 Faculty of Health Sciences, School of Dentistry, University of Copenhagen, Denmark

3 Department of Quality and Safety of Oral Health Care, Radboud University, Nijmegen, the Netherlands

4 Translational Health Economics Group (THE Group)/Department of Conservative Dentistry, Heidelberg University, Heidelberg, Germany

Introduction

Using claims data for research has great potential, because these data include large numbers of observations and have a high reliability [1], [2], [3], [4]. Therefore, such data are used, for example, to analyze treatment provision [5], [6] or health service utilization [7]. Using claims data for research is well established in the United States [8], [9] and in single European countries [10], [11]. However, multi-national approaches scarce.

A novel approach aimed to use claims data on European level for dental care research, by assessing the quality of processes of dental health care services and to optimize oral health towards prevention. The EU-funded project ADVOCATE ("Added Value for Oral Care") analyzes claims data from health insurance providers, health funds and health authorities in six European countries: Denmark, Germany, Hungary, Ireland, the Netherlands, and the United Kingdom [12]. The data were acquired from the data owners into a central data repository. Oral health measures for dental care were defined by a group of experts and were analyzed to determine which national characteristics had a positive influence on these indicators [13].

The measures refer to topics regarding the access to dental care, symptoms and diagnosis, health behaviors, oral prevention and patient perception. Examples for oral health measures are the number of specific treatments, for example periodontal examination, root canal treatment and tooth extraction in a specific time period. The aim was and is to recommend successful strategies to other countries.

Claims data are stored in databases of a data owner and specific data sets had to be generated for the ADVOCATE project. To what extent and under which conditions the data can be used for research is determined by each data owner [14]. The conditions refer both to country-specific data protection and privacy acts and internal policies [15], [16]. The variety of characteristics in these regulations require different approaches to access, transfer, and storage of the data.

In this paper, the generic steps, which are necessary to establish an individual data acquisition process, are described and formalized in a process model, to provide researchers with a methodological framework. Thus, the aim of this paper is to describe a standardized approach for acquiring international administrative routine data for research, based on the experiences gained during the ADVOCATE project.

Methods

A data acquisition process model was developed in an evolutionary process by adapting and tailoring the methodology of agile modeling [17]. The process followed the guidelines and recommendations of the "Good Practice of Secondary Data Analysis" (GPS), which provides guiding principles for conducting studies based on secondary data analyses [18]. GPS was systematically analyzed and necessary activities were determined. Information and data flow were formalized according to the Business Process Modeling Notation (BPMN) [19] using Microsoft Visio 2013 [20].

The next step was to determine specific activities and implementations of the general process for the ADVOCATE project. Therefore, a qualitative content analysis of data usage agreements (DUAs), which were consented with the data owners, was performed [21]. A coding scheme for data usage, data protection, and privacy aspects was developed. The code structure was developed in a mixed deductive-inductive approach [22]. Deductively created codes represent general aspects of data access, transfer and usage. Inductively created codes represent implementations in ADVOCATE. The relevant text passages in the DUAs that contained statements on specific implementations for the data acquisition process were then identified. The DUAs were examined and systematically coded by three independent reviewers using the software package MAXQDA 11 [23].

Results

GPS analysis

GPS was analyzed with a focus on practical aspects for the data acquisition process. Because GPS contains recommendations not only for acquiring secondary data, not all of the recommendations were implemented and certain activities were either added or were deemed to be irrelevant. Therefore, all guidelines were included, except for those, which focused on the planning of the study and which are not related to data acquisition (guideline 1, 2, 4). Furthermore, guidelines that are related to later stages than data acquisition (guideline 7, 10, 11) were excluded. The recommended plausibility checks in guideline 6.4 were extended towards the three-level quality and validity audit according to Horenkamp-Sonntag [24].

The analysis of GPS resulted in a four-phase data acquisition process that describes the general workflow: First, data access and usage are prepared, which involves the formal definition of general conditions. These conditions are subsequently tested by sharing a data sample. A quality and validity audit is performed next to check data quality of the data sample and the compliance of the data acquisition process with the regulations. After optimizing the process model, full-scale data access takes place to exchange all data as contracted.

Each phase contains several activities that need to be agreed with each data owner individually. The particular activities and recommendations for their implementation were deduced from the results of the qualitative content analysis and are described in the following sections.

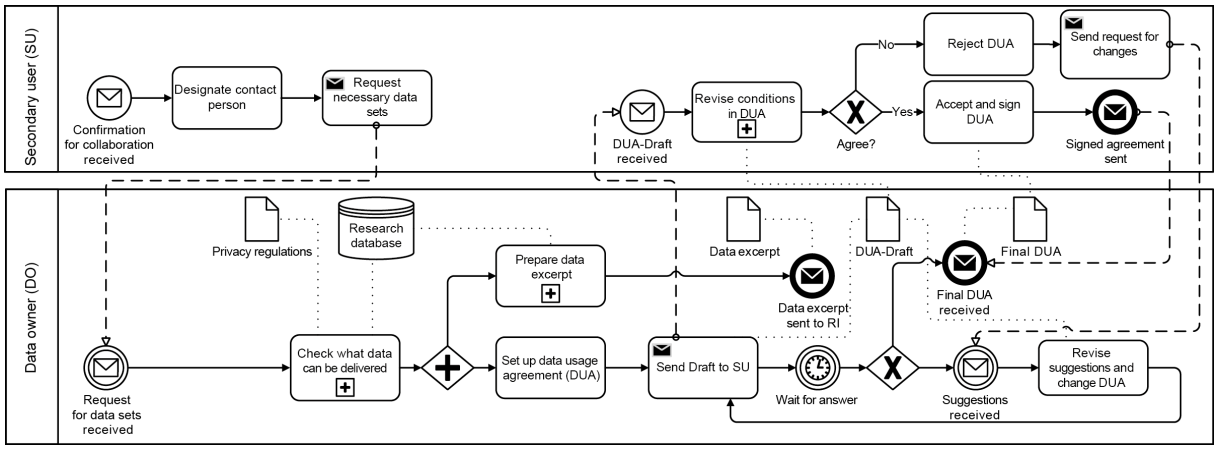

Figure 1: The first phase of the data acquisition process involves the preparation of the data acquisition.

Preparation of data access and usage

At the beginning of the project, the prerequisites to initiate and coordinate the data acquisition process are determined. In this paper, the terms "data owner" (DO) for the licensor of the data for secondary use, and "secondary user" (SU) for the licensee are used. This is in accordance with the recommendations of GPS [18]. Figure 1 depicts the sequentially performed tasks and the information exchange between the data owner and the secondary user. After the data owners confirm their willingness to collaborate with the secondary user, coordination on the specifications of the data that is to be provided follows.

Data characteristics (columns, profiles, and number of records) and data quality (individual level or aggregated) should be clearly specified in a DUA together with additional aspects concerning data usage, access, and storage. This includes the specification of underlying legal basis and individual policies of the data owner. Furthermore, all persons who will gain access to the data should be named and different roles and affiliated responsibilities should be defined. The DUA should also specify the location where data analyses take place. If the data are allowed to be transferred via the Internet, it should be stated how the data must be prepared (anonymized, pseudonymized, encrypted) and protected. In cases where legal regulations do not allow the transfer of data to the secondary user, it might be stated that the analyses can be conducted in the premises of the data owner. If data need to be archived further for retention reasons, it must be stipulated which party is responsible and how long the data will be archived. It should also be specified, whether the usage of the data is only allowed for a specific research project or if further analyses are allowed.

The draft of the DUA can be provided by the data owner or it can be developed by both parties. The secondary user has to check the conditions, in consultation with their legal advisers prior to signing the agreement.

Sharing a data excerpt

After the prerequisites are regulated, the consented approach should be tested with a data sample (Figure 2). The sample should be an excerpt of the full-scale data set and prepared in the same format (columns, profiles, variables, and value ranges). There are two different approaches in this phase:

1. Data access is only provided in the premises of the data owner. In this case, the secondary user receives the excerpt to develop data analysis routines. A data processor accesses the database in the premises of the data owner later, executes the routines and transfers the results to the secondary user's premises.
2. Full-scale data sets are transferred to the secondary user for analyses. In this case, the excerpt is used to develop and test the data transfer approach. The transfer of the excerpt should be performed in the same way as that for the full-scale transfer, in order to detect problems or deviation from the regulations. In addition to the data sample, the data owner should provide a description of the data, including the description of the variables, attributes, and treatment codes of services.

Quality and validity audit

As claims data are not originally collected for research purposes, their quality and validity, especially regarding completeness and plausibility, must be checked in all phases of the data acquisition process (Figure 3 and Figure 4). The audit should be performed by the secondary user, to ensure the data are suitable for the planned analyses. After the data excerpt has been transferred to the secondary user, the data quality and validity can be assessed regarding technical and project-related aspects. The technical audit involves the evaluation of completeness and plausibility of the parameters. Besides the content, the suitability of data transfer and storage approaches should be checked before full-scale data access and data analyses.

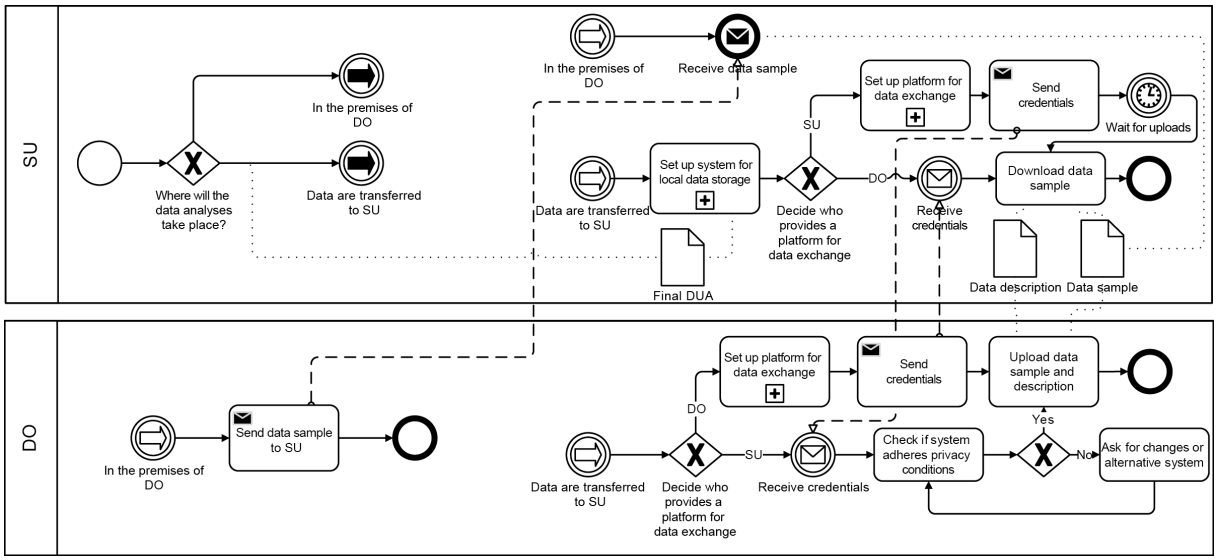

Figure 2: The second phase of the data acquisition process involves the exchange of a data sample.

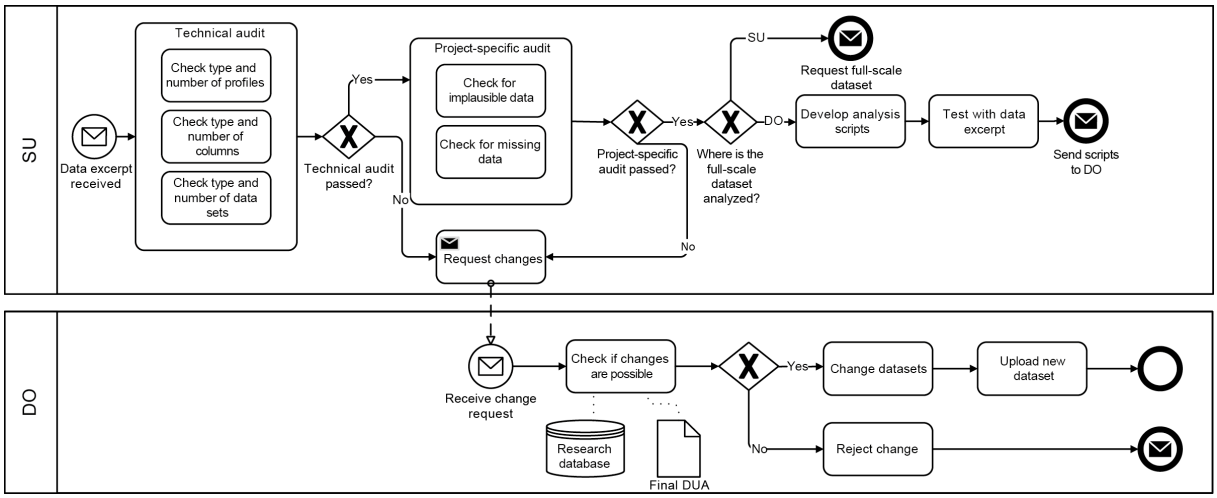

Figure 3: The third phase of the data acquisition process involves the performance of the technical and project-related quality audit.

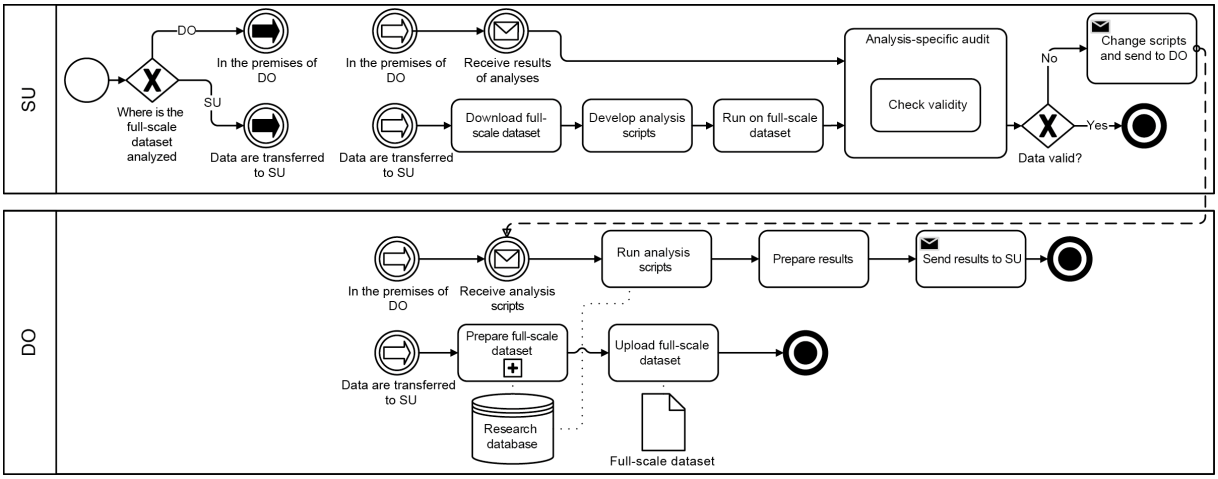

Figure 4: The fourth phase of the data acquisition process involves the performance of the full-scale data access and the analysis-related quality audit.

The project-related audit involves the evaluation of the suitability of the delivered profiles, columns, and records to perform the planned analyses.
Incorrect data should be identified and it should be checked if missing or incomplete data might be a problem. An upper limit of what is considered acceptable should be agreed. In case of inconsistencies, data owners should provide assistance and if needed, to deliver new corrected data.

Full-scale data access

The full-scale data sets contain all records that are consented with the data owners. It depends on whether full-scale analyses take place at the secondary user or at the premises of the data owner (Figure 4). After scientific analyses have been performed, the plausibility of the results has to be evaluated during the analysis-specific audit.

Implementation of the data acquisition process in the ADVOCATE project

The ADVOCATE project involved the acquisition of claims data from eight health insurance providers, health funds or health authorities from Denmark, Germany, Hungary, Ireland, the Netherlands, and the United Kingdom. To calculate the oral health care measures, the data owners were asked to provide data containing dental procedures, oral health outcomes, and socio-economic information. The data acquisition process was influenced by the fact that some of the data owners neither had experience in sharing their data for research nor had any predefined usage rules for them. Therefore, the four-phase process was developed to perform the data acquisition with each data owner in a comparable and reproducible way.

Preparation of data access and usage

The data owners had already agreed to collaborate before the data acquisition process started. The necessary requirements and regulations regarding data protection and privacy were then determined with seven of the data owners and DUAs were concluded. Only one data owner did not require to determine a DUA. The agreements mostly applied country-specific legal regulations and internal policies of the data owners. Therefore, an individual agreement was negotiated with each data owner instead of consenting the same agreement with all data owners. After DUAs were concluded, a qualitative content analysis was performed. The code structure covers characteristics concerning the provision of data by the data owner, data access and transfer, and scientific usage of the data. The entire code structure is depicted in Figure 5.

1. *Provision of data by the data owner:* DUAs determined preparatory measures that the data owners used before data were shared for research. Either data were anonymized by removing specific attributes and aggregating values to make them indistinguishable or data were pseudonymized. DUAs also determined the content of the data.
2. *Data access and transfer:* The specific regulations influenced data access and transfer. Seven data owners permitted data transfer to the secondary user via Internet, if data protection measures were implemented during transfer and storage at the receiving institute's premises. Only German regulations did not allow the transfer of data in general and Hungarian regulations did not allow the transfer of personal data.
3. *Scientific usage of the data by the secondary user:* All DUAs only allowed using the data for the analyses in the ADVOCATE project and restricted further usage. Most of the data owners required that the data have to be returned or destroyed after the end of the ADVOCATE project or after a specific period. However, the data owners then committed to keep the original data sets for a specific period, in case the secondary user needs them again later. Inter-coder reliability was assessed, to check if both reviewers had the same understanding of the conditions mentioned in the documents. From the results of the coding, specific implementations were derived and developed afterwards.

Sharing a data sample

An infrastructure for secure data handling and analysis was set up, using the collaborative data analysis platform AnalytiXagility [25], and individualized according to the requirements of the respective data owners. The analysis of the DUAs revealed that some policies of the data owners do not support the use of the collaborative data analysis platform. Furthermore, some data owners provided their own system or shared anonymized data via email.
Four of the data owners in the ADVOCATE project had initially shared a data sample. The data owners from Denmark, Germany, Ireland, and the Netherlands shared data excerpts from the full-scale data that contain between 100 and 500,000 cases. Other data owners shared an overview (descriptive statistics) of their data. Additionally, every data owner shared a description of the data, including the description of the variables, attributes, and treatment codes of dental services.

Quality and validity audit

During the technical audit, completeness and types of columns, records, and profiles were evaluated. For example, one data set contained many records with edentulous patients, which appeared implausible. Hence, the data owners were contacted and asked for correction.

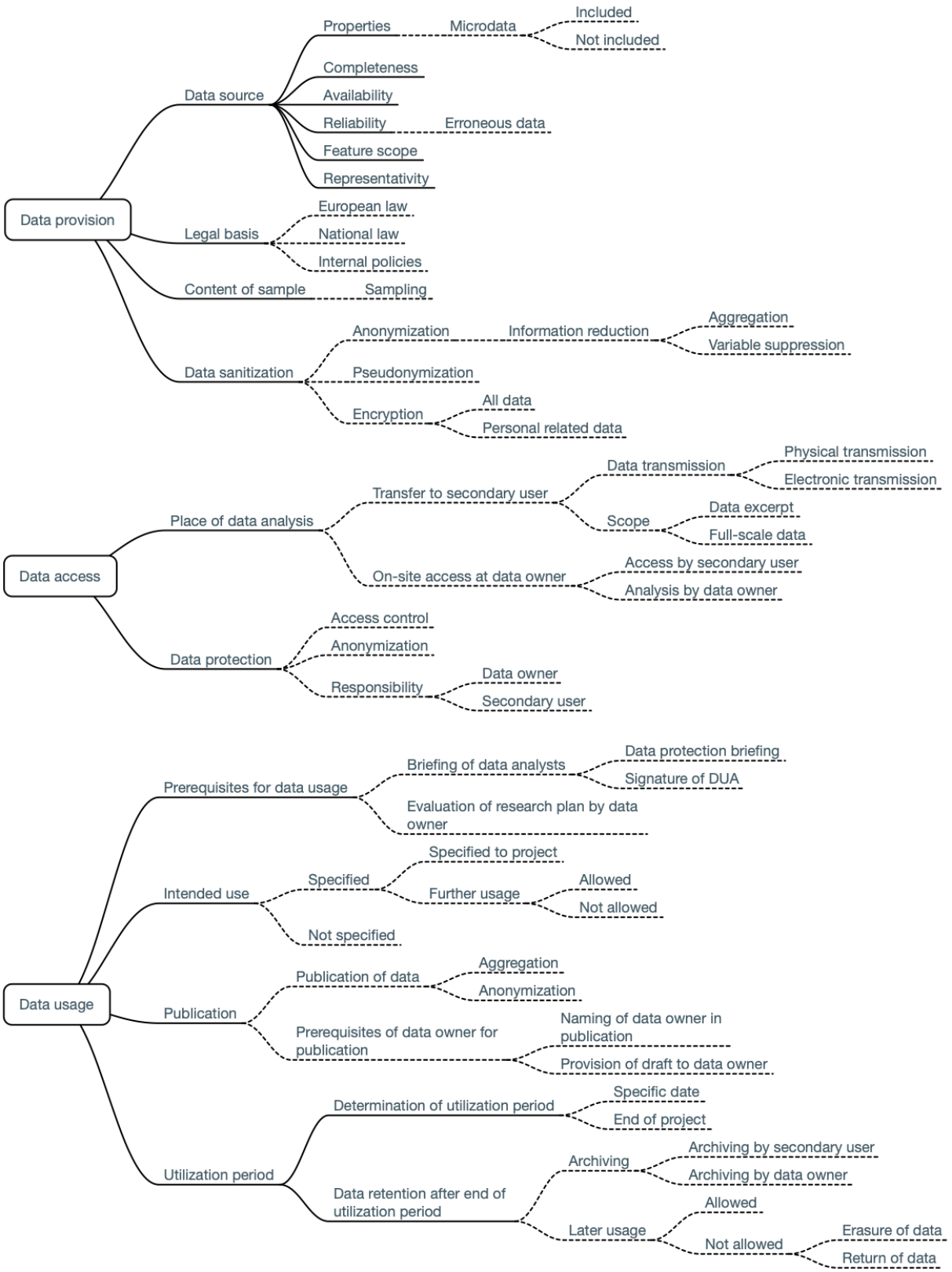

Figure 5: The coding scheme of the qualitative content analysis. The codes on solid lines were created deductively, representing general aspects of the data acquisition process. The codes on dotted lines were created inductively, representing implementations in ADVOCATE.

Table 1: Overview of administrative data bases and characteristics of available data

Country	Defined data sharing process	Available data	Data access	Data sanitization
Denmark	No	72,155,539 claims	Transfer of full-scale data	Anonymization
Germany	Yes	Sample of 100,000 claims	Transfer of data sample On-site access to full-scale data	Anonymization
Hungary	No	Aggregated data	Transfer of aggregated data	Anonymization Aggregation
Ireland (private)	No	411,797 claims	Transfer of full-scale data	Anonymization
Ireland (public)	No	Aggregated data based on 1,925,760 claims	Transfer of aggregated data	Anonymization Aggregation
Netherlands	Yes	89,004,606 claims	Transfer of full-scale data	Anonymization
United Kingdom	Yes	Aggregated data based on 40,018,073 FP17 forms	Transfer of aggregated data	Anonymization Aggregation
Scotland	Yes	4,369,961 claims	Transfer of full-scale data	Anonymization Aggregation

During the project-related audit, it was evaluated whether the planned analyses were possible with the provided data. In two cases, the data owners provided data that were too aggregated for the analyses. In these cases, the data owners were asked for less aggregated data, which was not possible due to privacy reasons.
Furthermore, in cooperation with dentists, it was evaluated whether the underlying treatment codes in the claims data were suitable to realize the planned analyses and whether they have the same meaning in the different reimbursement systems.

Full-scale data access

When the quality of the sample was adequate, the analyses of the full-scale data assets were performed (Figure 4). Seven data owners sent the full-scale data sets to the secondary user. In the case of the German data owner, analysis scripts were developed using the data excerpt and executed on the full-scale data in the premises of the German data owner. Only aggregated results were exported. Some of the data sets contained millions of records from 2010 to date. Table 1 shows the characteristics of available data.

Discussion

The use of claims data for research has great potential and can be used to evaluate effects of policy interventions on oral health outcomes [26]. However, accessing and using such data is challenging, because it often depends on individual coordination processes with the data owners [2]. Therefore, legal, ethical and technical prerequisites must be determined and agreed with each data owner separately. If multiple stakeholders are involved, researchers are confronted with additional challenges. Due to the lack of uniformity of legal regulations, it will rarely be possible to find a solution that fits all requirements of all data owners. A harmonized legal framework across and within Europe might facilitate the usage of claims data for research and increase data quality [27].
However, access to claims data will still be based on individual agreements with data owners. A heterogeneous data acquisition approach also poses a risk to data quality and the feasibility of analyses, because if data are provided on different quality levels, they might not be comparable or linkage with other data sources might not be possible [28], [29].
This paper has described how a generic approach has been developed, within the ADVOCATE project, to facilitate the data acquisition process and implement the individual requirements of each data owner.
There are comprehensive models for data acquisition and data quality assessment. For example, the controlled data sharing model of Prasser [30] describes a model, where secondary users can access data from a data owner via controlled data access environments. De Lusignan [31] describes six key concepts that researchers can use to assess the quality and utility of acquired routine data from data origin to curation and publication. Unlike these existing data acquisition and data quality assessment frameworks, the here presented approach was implemented pragmatically due to different regulations and different experience levels of the data owners in sharing data for research purposes. Whereas the Scottish health authorities provided a clearly defined application process for data usage [32], there were some data owners that neither had any experience in sharing their data for research nor had any predefined usage rules.
In this paper, it was shown that the data access in the ADVOCATE project was achieved by either accessing the data in the premises of the data owner or the data can

be transferred to the secondary user. Besides these two approaches, there can be more complex ways to exchange data for secondary usage [33], [34].
By performing quality and validity audits and feedback of the results to the data owners, it was possible to affect the scope or quality of the delivered data in some cases. For example, some data owners delivered additional data when they were requested. In other cases, it was not possible to deliver different data, due to privacy reasons. The experiences made during the ADVOCATE project show that a prescribed approach from the data owner is easier to implement, whereas the lack of experience on the side of the data owner gives an opportunity to tailor the data acquisition to the needs of the research project and may therefore become more flexible.
The process, described in this paper, can help researchers to overcome challenges with acquiring claims data from multiple data owners by providing a framework with general actions and processes. It also helps them to increase the efficiency and effectiveness of their processes. By defining the generic four-phase process in general and keeping the implementation open, it is possible to accommodate the individual requirements of all data owners.
The approach is limited insofar, as it only involves activities for data collection. It could be improved by extending it to include aspects of data harmonization to further increase data quality and to reduce risk of inability to perform data linkage. Also, it is solely based on experiences gained during the ADVOCATE project.
A standardized approach that supports researchers from the beginning of the process of data acquisition to data integration can help to achieve the highest possible improvement of knowledge through safe and efficient use of available data sources.

Notes

Funding

The ADVOCATE project has received funding from the European Commission's Horizon 2020 research and innovation program under grant agreement 635183 (http://www.advocateoralhealth.com/).

Acknowledgement

We are grateful to Professor Helen Whelton, coordinator of the ADVOCATE Consortium, for her pivotal role in bringing people together and initiating the writing of the grant of the ADVOCATE project and for her comments and suggestions in the final proof-reading.

We thank the contributors to the ADVOCATE project:

- the ADVOCATE Scientific Advisory Board – Stephen Birch, Martin Chalkley, Roger Ellwood, Ekatarina Fabrikant, Jeffery Fellows, Christopher Fox, Frank Fox, Dympna Kavanagh, John Lavis, Roger Matthews, Mariano Sanz, Paula Vassalo, and Sandra White;
- the ADVOCATE General Assembly – Renske van der Kaden, Onno van der Galien, Lisa Bøge Christensen, Gail Douglas, Kenneth Eaton, Gerard Gavin, Geert van der Heijden, Jochen Walker, Stefan Listl, Gabor Nagy, Karen O'Hanlon, Andrew Taylor, Helen Whelton, Noel Woods;
- the ADVOCATE Ethics Advisory Board – Mary Donnelly, Eckart Feifel, Jon Fistein, Joakim Lilholt, Orsolya Peter, Elmar Reich, Evert-Ben van Veen, Agnes Zana;
- the ADVOCATE project coordinator Anita Blakeston; and
- the coworkers of the ADVOCATE project – Susan Ahern, Hanny Alwicher, Fatiha Baâdoudi, Robert Bryce, Julia Csikar, Frank Gabel, Melinda Madlena, Neal Maskrey, Emma McDonald, Kasper Rosing, Tanja Schamma, Anna-Lena Trescher, Karen Vinall-Collier, Elaine Ward, and Evangelos Zormpas.

Competing interestes

The authors declare that they have no competing interests.

References

1. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005 Apr;58(4):323-37. DOI: 10.1016/j.jclinepi.2004.10.012
2. Lane J, Schur C. Balancing access to health data and privacy: a review of the issues and approaches for the future. Health Serv Res. 2010 Oct;45(5 Pt 2):1456-67. DOI: 10.1111/j.1475-6773.2010.01141.x
3. Smeets HM, de Wit NJ, Hoes AW. Routine health insurance data for scientific research: potential and limitations of the Agis Health Database. J Clin Epidemiol. 2011 Apr;64(4):424-30. DOI: 10.1016/j.jclinepi.2010.04.023
4. Leake JL, Werneck RI. The Use of Administrative Databases to Assess Oral Health Care. J Public Health Dent. 2005 Mar;65(1):21-35. DOI: 10.1111/j.1752-7325.2005.tb02783.x
5. Guiney H, Felicia P, Whelton H, Woods N. Analysis of a payments database reveals trends in dental treatment provision. J Dent Res. 2013 Jul;92(7 Suppl):63S-9S. DOI: 10.1177/0022034513484327
6. Suominen-Taipale AL, Widström E. A longitudinal study of young Finnish adults' use of subsidized, private sector dental care, 1986-1997. Community Dent Oral Epidemiol. 2000 Oct;28(5):365-72.
7. Listl S, Chalkley M. Provider payment bares teeth: dentist reimbursement and the use of check-up examinations. Soc Sci Med. 2014 Jun;111:110-6. DOI: 10.1016/j.socscimed.2014.04.020
8. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, Hsia DC. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 1992 Feb;82(2):243-8.

9. Dasanayake AP, Li Y, Wadhawan S, Kirk K, Bronstein J, Childers NK. Disparities in dental service utilization among Alabama Medicaid children. Community Dent Oral Epidemiol. 2002 Oct;30(5):369-76.
10. Rosing K, Hede B, Christensen LB. A register-based study of variations in services received among dental care attenders. Acta Odontol Scand. 2016;74(1):14-35. DOI: 10.3109/00016357.2015.1034771
11. Raedel M, Hartmann A, Bohm S, Priess HW, Samietz S, Konstantinidis I, Walter MH. Four-year outcomes of restored posterior tooth surfaces-a massive data analysis. Clin Oral Investig. 2017 Dec;21(9):2819-25. DOI: 10.1007/s00784-017-2084-4
12. Leggett H, Duijster D, Douglas GVA, Eaton K, van der Heijden GJMG, O'Hanlon K, Whelton S, Listl S; on Behalf of the ADVOCATE Consortium. Toward More Patient-Centered and Prevention-Oriented Oral Health Care: The ADVOCATE Project. JDR Clin Trans Res. 2017;2(1):5-9. DOI: 10.1177/2380084416668167
13. Baâdoudi F, Trescher A, Duijster D, Maskrey N, Gabel F, van der Heijden GJ, Listl S. A Consensus-Based Set of Measures for Oral Health Care. J Dent Res. 2017 Apr;96(8):881-7. DOI: 10.1177/0022034517702331
14. Nimptsch U, Bestmann A, Erhart M, Dudey S, Marx Y, Saam J, Schopen M, Schröder H, Swart E. Zugang zu Routinedaten. In: Swart E, Ihle P, Gothe H, Matusiewicz D, editors. Routinedaten im Gesundheitswesen: Handbuch Sekundärdatenanalyse: Grundlagen, Methoden. 2nd ed. Bern: Huber; 2014. p. 270-90.
15. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. Health Serv Res. 2010 Oct;45(5 Pt 2):1468-88. DOI: 10.1111/j.1475-6773.2010.01142.x
16. Lee LM, Gostin LO. Ethical collection, storage, and use of public health data: a proposal for a national privacy protection. JAMA. 2009 Jul;302(1):82-4. DOI: 10.1001/jama.2009.958
17. Ambler SW. Agile modeling: effective practices for eXtreme programming and the unified process. 1st ed. New York: Wiley; 2002.
18. Swart E, Gothe H, Geyer S, Jaunzeme J, Maier B, Grobe TG, Ihle P; German Society for Social Medicine and Prevention; German Society for Epidemiology. Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen [Good Practice of Secondary Data Analysis (GPS): guidelines and recommendations]. Gesundheitswesen. 2015 Feb;77(2):120-6. DOI: 10.1055/s-0034-1396815
19. Object Management Group. About the business process model and notation specification version 2.0. [cited 2017 Aug 31]. Available from: www.omg.org/spec/BPMN/2.0/
20. Microsoft Corporation. Microsoft Visio 2013 [software]. Version 15. 2012 [cited 2017 Aug 25]. Available from: https://products.office.com/en-us/previous-versions/microsoft-visio-2013
21. Haux C, Ganzinger M, Gabel F, Trescher A, Whelton H, van der Heijden GJMG, et al. Qualitative content analysis of data usage agreements for data aggregation in a European wide oral health care project. In: Grill E, Müller M, Mansmann U, editors. Health – exploring complexity: an interdisciplinary systems approach HEC2016; 2016 Aug 28 - Sep 2; Munich, Germany. Dordrecht: Springer Science+Business Media; 2016. p. S75.
22. Mayring P. Qualitative Content Analysis. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research. 2000 Jun;1(2). Available from: https://utsc.utoronto.ca/~kmacd/IDSC10/Readings/Readings/text%20analysis/CA.pdf
23. VERBI GmbH. MAXQDA [software]. Version 11.1.2. 2015 [cited 2017 Aug 25]. Available from: https://www.maxqda.com
24. Horenkamp-Sonntag D, Linder R, Wenzel F, Gerste B, Ihle PD. Datenvalidierung: Prüfung der Datenqualität und Validität von GKV-Routinedaten. In: Swart E, Ihle P, Gothe H, Matusiewicz D, editors. Routinedaten im Gesundheitswesen: Handbuch Sekundärdatenanalyse: Grundlagen, Methoden. 2nd ed. Bern: Huber; 2014. p. 314-29
25. Aridhia. The AnalytiXagility Approach. 2017 [cited 2017 Apr 3]. Available from: http://www.aridhia.com/platform/analytixagility/
26. Listl S, Jürges H, Watt RG. Causal inference from observational data. Community Dent Oral Epidemiol. 2016 10;44(5):409-15. DOI: 10.1111/cdoe.12231
27. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, Bernal-Delgado E, Blomberg N, Bock C, Conesa A, Del Signore S, Delogne C, Devilee P, Di Meglio A, Eijkemans M, Flicek P, Graf N, Grimm V, Guchelaar HJ, Guo YK, Gut IG, Hanbury A, Hanif S, Hilgers RD, Honrado Á, Hose DR, Houwing-Duistermaat J, Hubbard T, Janacek SH, Karanikas H, Kievits T, Kohler M, Kremer A, Lanfear J, Lengauer T, Maes E, Meert T, Müller W, Nickel D, Oledzki P, Pedersen B, Petkovic M, Pliakos K, Rattray M, I Màs JR, Schneider R, Sengstag T, Serra-Picamal X, Spek W, Vaas LA, van Batenburg O, Vandelaer M, Varnai P, Villoslada P, Vizcaíno JA, Wubbe JP, Zanetti G. Making sense of big data in health research: Towards an EU action plan. Genome Med. 2016 Jun;8(1):71. DOI: 10.1186/s13073-016-0323-y
28. Verschuuren M, Badeyan G, Carnicero J, Gissler M, Asciak RP, Sakkeus L, Stenbeck M, Devillé W; Work Group on Confidentiality and Data Protection of the Network of Competent Authorities of the Health Information and Knowledge Strand of EU Public Health Programme 2003-08. The European data protection legislation and its consequences for public health monitoring: a plea for action. Eur J Public Health. 2008 Dec;18(6):550-1. DOI: 10.1093/eurpub/ckn014
29. Eichler GS, Cochin E, Han J, Hu S, Vaughan TE, Wicks P, Barr C, Oevenport J. Exploring Concordance of Patient-Reported Information on PatientsLikeMe and Medical Claims Data at the Patient Level. J Med Internet Res. 2016;18:e110. DOI: 10.2196/jmir.5130
30. Prasser F, Kohlmayer F, Spengler H, Kuhn KA. A Scalable and Pragmatic Method for the Safe Sharing of High-Quality Health Data. IEEE J Biomed Health Inform. 2018 Mar;22(2):611-22. DOI: 10.1109/JBHI.2017.2676880
31. de Lusignan S, Liaw ST, Krause P, Curcin V, Vicente MT, Michalakidis G, Agreus L, Leysen P, Shaw N, Mendis K. Key concepts to assess the readiness of data for international research: data quality, lineage and provenance, extraction and processing errors, traceability, and curation. Contribution of the IMIA Primary Health Care Informatics Working Group. Yearb Med Inform. 2011;6:112-20.
32. The Scottish Government eHealth Division. Public Benefit and Privacy Panel for Health. 2017 [cited 2017 Apr 3]. Available from: http://www.informationgovernance.scot.nhs.uk/pbpphsc/
33. Lablans M, Kadioglu D, Muscholl M, Ückert F. Exploiting Distributed, Heterogeneous and Sensitive Data Stocks while Maintaining the Owner's Data Sovereignty. Methods Inf Med. 2015;54(4):346-52. DOI: 10.3414/ME14-01-0137
34. Peters E, Maxemchuk N. A privacy-preserving distributed medical insurance claim clearinghouse & EHR application. In: Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017; 2017 Jul 17-19; Philadelphia, USA. Washington: IEEE; 2017. p. 70-76. DOI: 10.1109/CHASE.2017.62

Corresponding author:
Christian Haux
Institute for Medical Biometry and Informatics, Heidelberg University, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany
Christian.Haux@med.uni-heidelberg.de

Please cite as
Haux C, Rosing K, Knaup P, Listl S, Kalmus O. A process model for acquiring international administrative routine data for health services research. GMS Med Inform Biom Epidemiol. 2019;15(1):Doc04. DOI: 10.3205/mibe000198, URN: urn:nbn:de:0183-mibe0001985

This article is freely available from
http://www.egms.de/en/journals/mibe/2019-15/mibe000198.shtml

Published: *2019-05-20*