Amplifying Domain Expertise in Clinical Data Pipelines

Authors: Protiva Rahman, Ph.D, Arnab Nandi, Ph.D., Courtney Hebert, M.D., M.S.
Affiliation: The Ohio State University

Abstract

Digitization of health records has allowed the healthcare domain to adopt data-driven algorithms for decision support. There are multiple people involved in this process: a data engineer who processes and restructures the data, a data scientist who develops statistical models, and a domain expert who informs the design of the data pipeline and consumes its results for decision support. While there are multiple data interaction tools for data scientists, few exist to allow domain experts to interact with data meaningfully. Designing systems for domain experts requires careful thought because they have different needs and characteristics from other end-users. There should be an increased emphasis on the system to optimize the experts' interaction by directing them to high impact data tasks and reducing the total task completion time. We refer to this optimization as amplifying domain expertise. While there is active research in making machine learning models more explainable and usable, they focus on the final outputs of the model. However, in the clinical domain, expert involvement is needed at every pipeline step: curation, cleaning, and, analysis. To this end, we review literature from the database, human-computer information, and visualization communities to demonstrate challenges and solutions at each of the data pipeline stages. We then present a taxonomy of amplifying expertise, which can be applied when building systems for domain experts. This includes summarization, guidance, interaction, and acceleration. Finally, we demonstrate the use of our taxonomy with a case study.

Introduction

Recent advancements in data availability (e.g., digitization of health records) and deep neural networks [1] has led to the resurgence of artificial intelligence. This has served as a catalyst for data-driven decision making in many domains. However, for high-stakes applications, such as financial and healthcare domains, it is rare for domain experts to execute decisions solely based on artificial intelligence algorithms [2]. Domain experts in this context are individuals who are not necessarily trained in computational fields, but inform the design and are end-users of data-driven algorithms (e.g., healthcare providers, hospital administrators). Note that domain experts can have different levels of expertise in their specific domain (e.g., interns, residents, attendings), and we do not differentiate between these levels in this work. While the role of experts has been studied in clinical decision support (CDS), we find a gap in their involvement in the data analysis pipeline, which we focus on.

Domain expert involvement remains necessary in the healthcare domain, but this involvement brings significant challenges and implications for data-driven applications. Domain experts are expensive resources with limited time for these efforts and excessive reliance on domain expertise could potentially lead to systems that are overly customized and not reproducible or scalable. Because of these challenges, designing systems for them requires careful thought. To address these challenges, we present a framework for amplified intelligence which identifies the points in the process where expertise can be effectively leveraged. *Amplification of expertise* then refers to the process of automating redundant or inferable tasks, so that domain experts can focus their efforts on tasks that require domain

knowledge. This is a synergy between the domain expert and the system, which involves summarization of data and decisions, guidance towards insights, interaction by the domain expert, and acceleration of input (Figure 1).

Prior Work

There is active research on interactive and human-in-the-loop systems in many computer science sub-disciplines. The database and visualization communities have produced numerous tools [3-8] to aid data scientists with data wrangling and analysis. At the decision-making stage, the machine learning community has looked at making black-box models explainable [2, 9-12] while the human-computer interaction (HCI) community has been studying how difference in explainability affects decision making [13, 14]. Finally, the crowdsourcing community has concentrated on human-powered computation by optimizing tasks (e.g., simplifying tasks [15], minimizing number of questions [16, 17], optimizing workflows [18-20]). However, we focus on data-powered experts by amplifying expertise. While we draw from prior work, systems designed for healthcare domain experts require special consideration because they have characteristics that distinguish them from data scientists and crowd-workers.

Special Considerations in the Healthcare Domain

First, domain expert input is usually needed for data tasks that require experiential knowledge and judgment (such as medical diagnoses, forensic analysis [21]). The critical and subjective nature of these decisions necessitates transparency, both from the algorithm as well as domain experts. Hence, the system needs to summarize the impact of algorithmic or experts' manipulation of the data [22]. Second, due to their specialized training, domain experts' time is expensive and limited[23, 218]. This constraint makes it imperative that we build tools that guide them to insights while reducing physical and cognitive effort [24]. Third, since domain experts are trained in non-computational fields, systems designed for them should provide high-level interaction capabilities. This is referred to as *editable shared representations* between computers and humans [25]. Examples of this include natural language interfaces and form-based input [26]. Finally, domain experts are highly trained individuals, which allows systems to accelerate their input by using domain-specific assumptions and ontologies [27, 28]. Keeping these factors in mind, expertise amplification involves summarization, guidance, interaction, and acceleration (Figure 1). We will explore each of these in detail in the following sections.
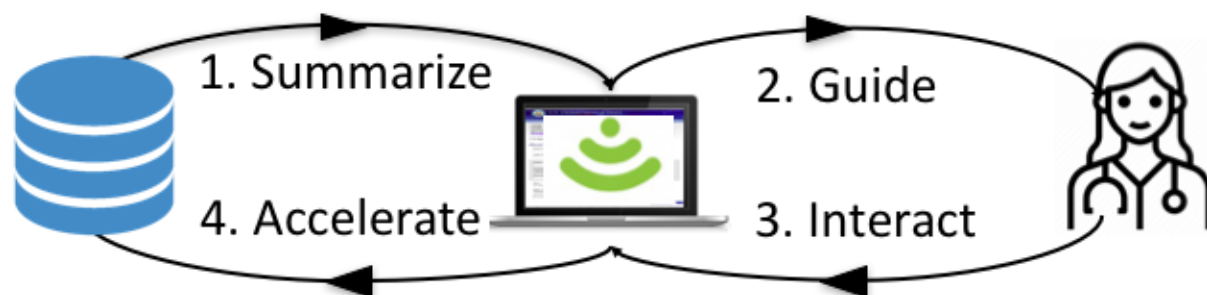

Figure 1: Domain Expertise Amplification

The Data Pipeline

There are opportunities to amplify expertise at all stages of the pipeline. The data pipeline refers to the different stages that the data need to go through before they can provide decision support. It can roughly be broken into three stages: curation, cleaning, and analysis. Tools at the end of the pipeline have only looked at explaining models but not at

amplification. In contrast, tools at earlier pipeline stages have been designed mainly for data scientists and not for experts. Domain experts, however, are involved at every stage of the pipeline [26-30], especially in clinical research settings where datasets contain specialized information. Thus, there is a need to amplify domain expertise throughout the pipeline. In this work, we provide examples from the informatics literature to highlight the need for expert involvement at each pipeline step. We then review literature from the database, HCI, and visualization communities about challenges and current approaches at different stages. Based on our review, we present a novel taxonomy for amplifying domain expertise and demonstrate its use with a case-study in empiric antibiotic prediction. Our review can serve as a guide to new clinical research projects, and our taxonomy can be applied when designing systems for experts, especially for low-budget projects when there are limited resources and availability of domain experts.

Challenges in the Data Pipeline

We organize this section to reflect the clinical data pipeline, which often involve the following steps: data is curated from the EHR data warehouse and annotated with external data sources, cleaned and validated, and analyzed. Multiple people are involved at various stages of the pipeline. The prevalent notion of the workflow is that a data engineer restructures, cleans, and sets up the infrastructure for the analysis for data, a data scientist then analyzes and models the data, which a software engineer implements into a decision support system. A domain expert then consumes the end-product to make decisions. However, in clinical settings, the specialized nature of the data often requires domain expert involvement at every step of the pipeline, which we will demonstrate in this section. Allowing domain experts to directly and efficiently interact with data removes the need for them to rely on a data engineer or data scientist, who can then respectively focus on infrastructure and model construction. Moreover, since domain experts are the stakeholders in the output of data pipelines, in our experience, they tend to be engaged users who want to interact with data. In this section, we motivate domain expert involvement with examples from the past five years of research presented at the American Medical Informatics Association (AMIA)'s annual symposiums. We review the computer science literature to identify current tools and opportunities for expertise amplification at the three stages of the data pipeline: data curation, data cleaning, and data analysis.

Data Curation
Curating data sets for analysis can be a laborious process that can involve combining multiple data sources and identifying relevant attributes. Data integration and data discovery address these respective problems.

Data Integration: Medical data pipelines often involve data that were collected for purposes other than answering the research question at hand. This usually implies that information is not captured in a manner fit for analysis [31, 32], with issues such as missing metadata information [33]. Moreover, in some situations such as rare disease studies, the cohort size is too small for analysis [34], while in other cases, external features such as air quality or drug components [35-38] might be needed. One possible solution to these data quality issues is to curate data from multiple institutions and external sources. However, the different data representations [34, 39] pose challenges in entity matching, metadata inference, and data integrity [40, 41]. Data integration aims to automatically resolve schema matching and entity matching problems during data curation. For biomedical datasets, integration can involve standardization by mapping to ontologies with controlled vocabularies [42-44]. While current approaches use deep learning for integration [45-49], generating a training corpus

and validating results require domain expert input. For example, Cui et al. require domain experts to validate data curation efforts for studying sudden death in epilepsy [34]. In another example, building an automatic concept annotator for standardizing biomedical literature [49] required experts to annotate different concepts [50-53] manually. Further, a domain expert will be able to catch inconsistencies or errors made by an automated integration tool much faster than a data engineer who is unfamiliar with the domain. Thus, there is a need for building interactive data integration tools for domain experts.

Data Discovery: Data discovery refers to the process of finding relevant attributes or cohort for analysis. This is especially true for multidisciplinary teams where the domain expert knows the disease definition but is not familiar with the database schema. At the same time, the data engineer can explore the schema but might not recognize that a field is relevant. Integrating data from multiple sources only exacerbates this problem. In the informatics community, DIVA[214] aids in cohort discovery by ingesting expert-defined constraints, while visual analytic systems[215, 216] such as CAVA provide an interactive interface. In the database community, Nargesian [54, 55] has looked at finding unionable (more data points) and joinable (more attributes) data for a given dataset. These algorithms are useful when trying to augment datasets with publicly available datasets such as MIMIC [56] or even for exploring a complex schema like the Unified Medical Language System (UMLS) [57]. In addition to using properties of the data to find possible matches, domain rules can be useful for identifying relevant data. This requires an interactive interface where domain experts can look at subsets of interest and iteratively join and filter the data [58] to find the required cohort. Recently, query logs have been used to design precision interfaces [59, 60], which customize the interface for the user's interactive task.

Data Cleaning

After curating relevant datasets, data still need to go through multiple preprocessing steps before they are analysis-ready. These include identifying and fixing incorrect data, data augmentation, and data transformation, all of which benefit from domain expert involvement.

Error Fixes: EHR data is known to be messy and have errors and missing values [62-64]. A typical data cleaning method is the use of rule-based systems that identify dirty data by detecting violations of user-specified rules or known functional dependencies [65-74]. But again, these systems do not optimize the expert's rule specification process. Crowdsourcing systems also have been used to correct values [18, 75], though they are not always an option due to data complexity or confidentiality. Another approach to identifying errrors and cleaning data is to augment it with external knowledge bases [76-78]. More recently, there have been many approaches [79-81] that use deep learning for automated data cleaning. Of note is Holoclean [80], which uses a statistical model to combine various data repair signals such as violation of integrity constraints, functional dependencies, and knowledge bases. While this achieves higher performance than using each method in isolation, there is scope for identifying which of the signals are performing the poorest or what additional information would help improve the system's performance. Identifying this information, incorporating domain knowledge, and presenting it succinctly to a domain expert remains an open problem.

Data Augmentation: While data entry errors [82] and missing information can be imputed by semi-automated methods, a more difficult problem is that of creating a gold standard for training data, which is referred to as data augmentation. Many healthcare applications require annotating training data, e.g., clinical text annotation [83-85], clinical decision support [86-88], identifying new terms for ontologies [89], index terms for articles [90],

disease-specific annotations [50, 91, 92]. However, very few applications focus on optimizing the domain expert's data augmentation effort, which is eventually crucial to model performance. A notable approach to this is the Snorkel system [93], which automates data augmentation by learning the labeling function, thus accelerating the domain expert's input. However, there are opportunities in making the initial labeling process more interactive since domain experts are required to write code. Further, the system does not provide feedback on how labels affect the data set or final model, which is crucial for building trust in medical pipelines. Examples of interactive solutions include Icarus [27] for augmenting microbiology data and Halpern et al.'s system[94] for annotating clinical anchors. Both these systems use an ontology to amplify domain expertise interactively.

Data Transformation: Other than fixing incorrect values and augmenting datasets, often, data need to be restructured (e.g., splitting values in a column, reformatting dates). Data wrangling has emerged as a separate field in the past decade because of data diversity. Potter's Wheel [95] was one of the first interactive data transformation systems. It allows the user to specify transforms which are encoded as constraints and used to detect errors. Building on this idea, systems such as Polaris [96] and Trifacta [4, 97] infer syntactic rules from user edits. Similarly, programming-by-example systems [98, 99] learn transformations from a set of input-output pairs. These techniques have informed the autofill function of Microsoft Excel. Since many domain experts employ Excel for data transformations and analysis [100], spreadsheet interfaces should consider incorporating domain knowledge.

Data Analysis

We now move to the final step of the pipeline. This includes initial exploratory analysis to identify attributes of interest, as well as explainability of models for decision making.

Data Exploration: During the exploration step, it is crucial for the domain expert to be able to directly interact with the data for effective hypothesis generation. However, domain experts often must go through a data engineer to execute the relevant query [101, 102] or extract information from unstructured notes[103]. The data are then validated by the domain expert through manual chart review, since data engineers without domain knowledge may apply naive filters which hide insights or find spurious correlations. To address these challenges, the informatics community has built tools to accelerate chart review [104] and allow interactive filtering and analysis [105, 106]. Finalizing an analysis dataset can then take multiple iterations of requests and validations between the domain expert and data engineer. In some cases, data engineers create custom dashboards for domain experts [107-109], but the latter is then limited to brushing and linking on the provided view. Mixed-initiative interfaces such as Tableau [96] and Dive [5] recommend visualizations based on statistical properties of the data, but do not use domain-specific ontologies which can enrich the domain experts' interaction and accelerate their workflow.

Visualizations are another technique, which can help the data exploration process. When used appropriately, visualizations can provide effective summaries and reveal patterns not immediately evident by statistical overviews [110]. Summaries reduce the cognitive load on domain experts during multidimensional data exploration, allowing them to drill-down to specific instances as needed [111]. While many visualization recommendation systems exist for analyzing numerical data [7, 112-114], visualizations in healthcare often include categorical and text data [115, 116-118]. As such, node-link diagrams are a common data representation and have been used for tracking family history [119], decision-making [22, 120], and identifying hidden variables [121]. Visual interfaces, thus amplify expertise by

summarizing data. However, they can be more powerful if they allow interaction, provide guidance by highlighting interesting regions for exploration [122], and accelerate workflows by extrapolating domain expert interactions based on properties of the data [22]. Thus, there is a need to provide domain experts with tools that allow for more sophisticated data interaction.

Explainability: Finally, we cannot discuss clinical pipelines without talking about explainability. The interpretability of rule-based systems has made them popular in a variety of clinical applications, including decision support [123, 124], antibiotic recommendation [125], updating annotations [126], and auditing [127]. Interpretability is essential because domain experts want a cause-and-effect relationship, based on which actionable decisions can be taken [62, 64, 128]. Further, healthcare providers may not use models they do not trust, and building trust requires providing context and explanations [2].
Current approaches in healthcare research use weights and activation of features to characterize attribute importance [129-131]. Ming et al.'s RuleMatrix [132] provides an alternate approach where a set of rules represents the deep learning model. The expert can explore various facets of each rule, such as data affected, distribution, errors, etc. In another example, Cai et al. [28] built a tool to help pathologists find similar images to aid in diagnoses. The tool allows domain experts to search for similar images and then interactively refine search results. It allows refinement by region (crop an image), refinement by concept (filter by extracted concepts from image embeddings), and refinement by example (select multiple images as examples). These refinement techniques are examples of acceleration, where interactions are interpolated to the entire dataset by learning general functions. Explainability is thus key for the adoption of deep learning models. While they have mainly been applied in the analysis stage of the pipeline, they are equally important when applying automated algorithms to curation and cleaning.

Table 1: Review of Current Approaches for each Data Pipeline Stage

	Theme	Current Solutions	Domain Expert Role
Data Curation	Data Integration	schema matching [134-139], interactive integration [140, 141], webtables integration [142-147], machine learning [45- 48]	Domain experts are needed to validate results of integration, and interactively correct automated methods, which can then update their algorithm.
	Data Discovery	attribute search [54, 55, 148, 149], interactive querying [58-60,214] visual analytics [215-216]	Domain expert feedback is needed to finalize the analysis dataset.
Data Cleaning	Error Fixes	rule-based [65- 73, 150] crowdbased [18, 75, 151, 152] knowledgebase [76-78] machine learning [79-81] functional dependency [74, 15-161]	Domain expert input can be used to identify and fix errors.
	Augmentation	machine learning [93, 162, 163] interactive [27, 94, 164-166]	Domain experts can augment missing data with domain-specific rules.
	Transformation	programming by example [99, 98] interactive rules [4, 95-97] foreign-key detection [149, 167-171]	Domain experts can restructure the data to make it semantically valid.
Data Analysis	Exploration	optimize performance [172-175] optimize insight [122, 176,177] provenance [178, 179] visualizations [5, 7, 112-114, 180-183, 217]	Domain experts interact with summaries and outliers to draw insight
	Explainable	systems [184, 185] visualizations [9, 12, 28, 132] empirical studies [10, 11, 13, 14]	Domain experts inform the model design to ensure explainability.

Amplifying domain experts' abilities in the analysis stage, therefore, requires interactive data systems using a combination of statistical algorithms and compelling visualizations. Moreover, these systems need to follow design-study principles [133]. System designers should consult domain experts during the design process and conduct empirical evaluation to ensure that the system effectively portrays the right information. Otherwise, the system can end up burdening and biasing the domain expert instead of helping [13, 97].

We have highlighted the need for domain expert involvement in the pipeline and described some of the challenges they encounter. While we have briefly expanded on some available solutions, Table 1 provides a more comprehensive list of references. Summarizing each technique is outside the scope of this paper, but it provides a guide to interested readers for further reading.

Taxonomy of Expertise Amplification

The previous section motivated the need for domain expert involvement throughout the clinical data pipeline. In all steps, domain expert involvement can improve automated methods, but must be done appropriately to ensure that the process remains robust and reproducible. Taking this into consideration, we propose a taxonomy that can be employed when designing systems to amplify expertise in the clinical pipeline. Domain expertise amplification by a system can broadly be categorized into four dimensions: summarization, guidance, interactivity, and acceleration, as shown in Figures 1 and 2. Thus, a system that wishes to amplify expertise should apply one or more of these dimensions. We demonstrate these categories with examples from computer science literature.
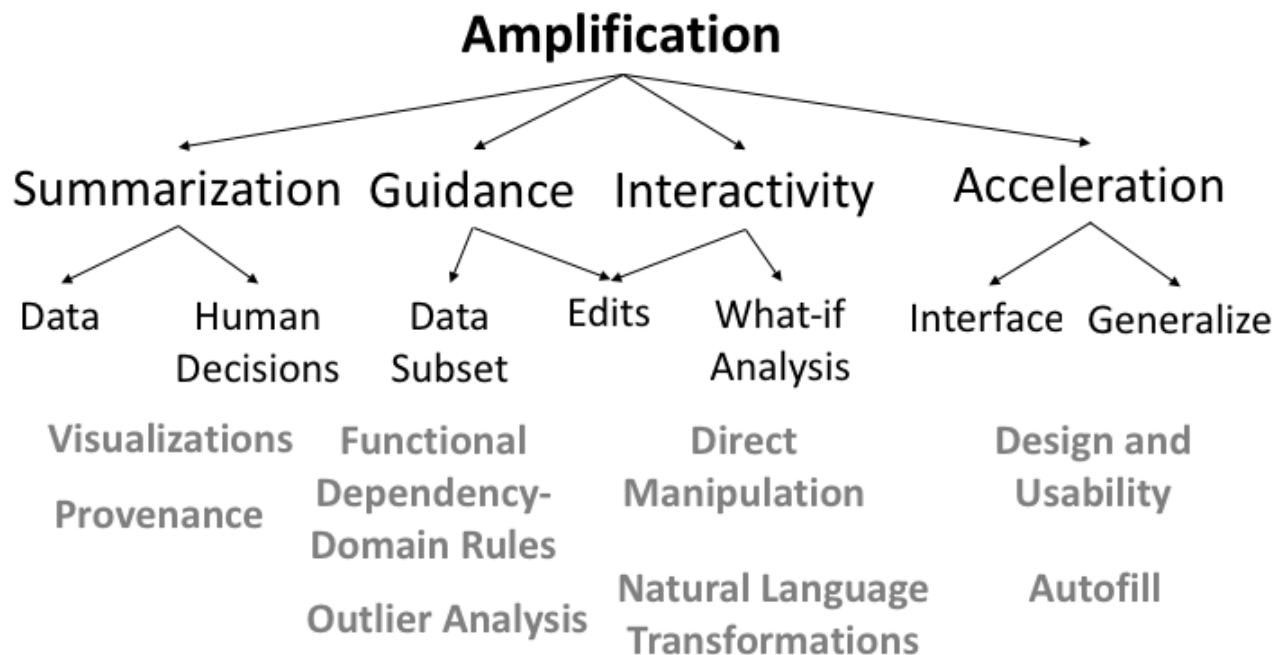
Figure 2: Taxonomy of Expertise Amplification: The first level shows the four dimensions which should be employed by a system for expertise amplification. The second level enumerates the sub-dimensions along which amplification can be done, while the fourth level in grey are tools that can be applied.

Summarization

The time constraints of experts, along with transparency requirements in the clinical domain, motivate the need for effective summaries of data and human decisions. While

data summaries are important for analysis, summaries of human decisions allow for improved explainability and reproducibility.

Data: An amplification system should summarize large and complex datasets so that experts can meaningfully consume them. This is relevant for identifying inconsistencies as well as for open-ended exploration during analysis. It can be overwhelming for an expert to go through large and wide tables. Amplification systems should, therefore, automatically summarize complex data [186]. While providing data samples [27, 72] and statistical summaries such as mean and variance can be useful for providing a bird's eye view, they are not always enough to reveal patterns [110]. In such cases, *visual summaries* can provide additional insight, as done by the CAVA system[215]. Multi-dimensional data can be visually summarized by presenting each dimension as a coordinated histogram with linked brushing and filtering [172].

Human Decisions: In addition to data, amplification systems need to summarize algorithmic and human decisions as well. This is because domain expert involvement is usually required in situations where it is necessary to have high-quality data [2, 21]. Hence amplification systems also require high transparency [184, 187]. To support algorithm transparency, amplification systems can show visual activation of features that led to the recommendation [9] or similar cases in the data that serve as evidence for the current recommendation [188]. Summarizing human decisions can involve expressing data transformations as natural language rules [4, 27] and as visual node-link diagrams [22]. Further, since summarized data provides an abstract or aggregate view, there is a need for data transparency, meaning that experts should be able to trace the individual data points, which contributed to the aggregate summary. This involves incorporating ideas from *provenance systems* such as Smoke [178] and Scorpion [179], which provide fast data lineage tracking. Finally, for each application, empirical studies are needed to see what and how information should be presented or summarized because too much transparency can overwhelm and negatively impact the expert [13].

Guidance

While summaries provide a global view of the data, goals of exploratory analysis include finding insights and data quality issues [186], which might require looking at a more detailed view. Systems can guide experts by navigating to informative subsets as well as by suggesting data transformations and edits.

Data Subset: Amplification systems should guide the expert's navigation to meaningful subsets. For example, SeeDB [112] automatically finds interesting visualizations. Given a query, it defines "interestingness" as the deviation of the query's resultset from a baseline dataset. Similarly**,** TPFlow [189] uses tensor decomposition to guide users in spatiotemporal exploration. For data cleaning, error detection algorithms such as Uguide [74] and DataProf [72] use *functional dependencies* and Armstrong samples, respectively, to find incorrect tuples for human validation, while Icarus [27] presents the expert with impactful subsets for data completion. Visual summary tools such as Profiler [180] use statistics to find data quality issues. When guiding users with visual summaries, it is important to pick optimal visual encodings for revealing the relevant insight or *outlier*. This can be informed by Correll et al.'s recent work [181], which empirically evaluated different visual encodings on their effectiveness at revealing data quality issues.

Edits: In addition to navigating datasets, amplification systems can also guide experts by suggesting data transformations to edit the data during the cleaning and preparation stage [4, 27, 99]. But even in this case, transparency is required. This is evidenced by the fact that

in empirical studies of Proactive Wrangler [97], users often ignored the suggested transformation but then manually performed the same one because the operation's semantics were unclear. Methods to aid in data transformation transparency include showing previews and transitions of the data changes [190] resulting from the transformation operation.

Interaction

Along with making system internals explainable [10], allowing experts to interact and modify data and the output of algorithms increases their trust in amplification systems [11]. For empiric antibiotic recommendation [191], this can involve allowing the healthcare provider to edit model features. Providing interaction comes at the cost of maintaining strict latency constraints since experts will expect to see the results of interaction almost immediately [133]. Techniques for maintaining interactive performance include sampling [192] and predictive prefetching [193]. Interaction modes can include data transformation suggestions and what-if analysis.

Data Transformation: The mode of interaction for data transformation in expertise amplification systems also needs to cater to their background and training. For example, transformations should be presented as *natural language* statements [4] as opposed to code snippets [93, 150]. While graphical user interfaces (GUIs) can decrease trust and control for system administrators [194], they are needed in amplification systems. Gestural query systems, such as GestureDB [58] and DBTouch [195], and *direct manipulation interfaces*, might be preferable to domain experts who are unfamiliar with SQL. Further, domain experts' affinity for spreadsheet tools [100] motivate designing systems with spreadsheet interfaces but advanced querying capabilities such as Dataspread [196] and Sieuferd [197].

What-if Analysis: To support collaborative decision making, amplification systems should allow for what-if analysis, where domain experts can apply or test different "decisions" and "assumptions" and see how it affects the dataset. Collaborative decision-making is important for consensus and conflict resolution. Domain experts are highly trained and experienced individuals in their fields, which affects how they interact with systems [198, 199]. Data pipeline tasks that require their input need them to apply knowledge from training and experience [27]. Such tasks inherently require judgment, which can be biased and can vary between and within domain experts [200]. To account for this bias, consensus from multiple experts is needed. However, unlike crowd workers, where differences in results can indicate bad actors entering random choices [18, 201, 202], in the case of domain experts, they reveal differing judgments. As such, automatic conflict resolution [203], such as majority voting, cannot be used since disagreements require expert discussion [22]. Collaboration is required for conflict resolution, and what-if analysis can speed up this process. Capturing and sharing metadata is also useful for collaboration [204-207].

Acceleration

Time constraints of domain experts necessitate the need to accelerate their input provision. This involves designing interfaces that aid the expert's task and building interactions that interpolate from edits to generalize to multiple data points.

Interface Design: Most experts use structured interfaces such as forms [208] or free text notes [209] for data entry or querying and spreadsheet interfaces for data exploration [100]. Following *user-centered interface* design and adhering to latency constraints is even more essential for these systems. Query interface layouts can be optimized by using statistical

properties of the data [210, 211, 212] and prior query logs [213, 60], while spreadsheet interfaces can be improved by incorporating higher expressibility [196, 197]. The Usher [211] system, an example of the former, uses a probabilistic model on prior input form data to optimize the form structure. This involves showing highly selective data attributes at the beginning of the form to reduce the complexity at later stages, thus reducing the scope of error and accelerating input provision.

Generalize: An advantage of building systems for domain experts is that domain-specific information can be used to accelerate their input. For example, Icarus [27] uses the organism and antibiotic hierarchy encoded as foreign-key relations in the database to generalize a single edit to a rule that fills in multiple cells, accelerating the data completion process. In another example, Cai et al. [28] allow domain experts to refine result sets with domain-specific concepts extracted from image embeddings.

Case Study

We now illustrate our taxonomy with a case study from a representative clinical data project: modeling empiric antibiotic treatment (Figure 3). We apply the four dimensions of amplification to the three stages of the pipeline. This is summarized in Table 2.
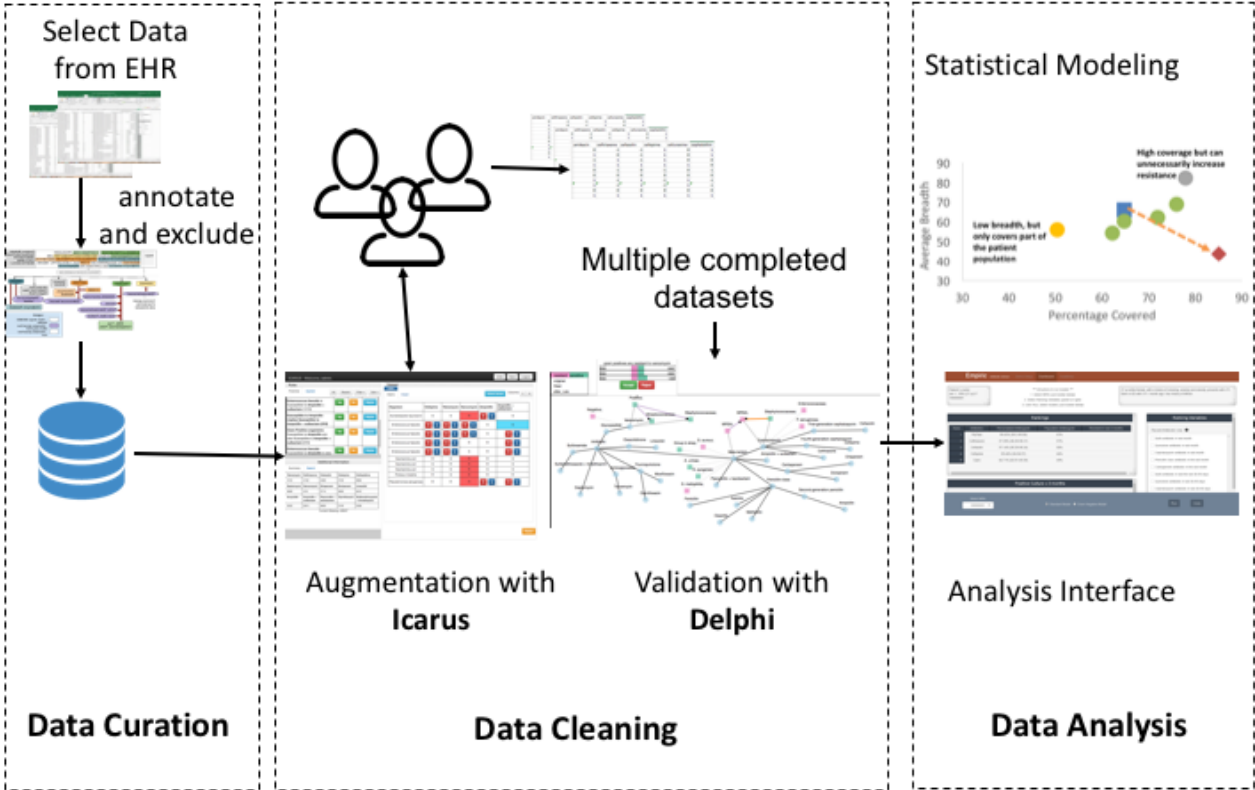

Figure 3: Data pipeline for Empiric Antibiotic Prediction

At the data curation level, our domain expert, Lucy, must provide the cohort definition along with variables of interest (e.g., demographics, comorbidities, allergies, etc.) to a data engineer, who pulls the relevant data from the EHR data warehouse. After the data pull, Lucy looks through the initial set and formulates additional exclusion rules to ensure it matches the clinical case definition. To implement these rules, the data engineer annotates the data with microbiology classification information the UMLS metathesaurus [57]. This process could be improved with an expertise amplification system. The system should *summarize* data by

showing the distribution of variables with linked brushing and filtering so that Lucy could see how the variable distributions are correlated. It could *guide* Lucy by suggesting correlated variables to the ones she selects. During validation of the cohort, Lucy should *interactively* be able to select datapoints to include. Finally, the system should be able to *accelerate* Lucy's validation by suggesting exclusion rules based on her interactions.

After the cohort is finalized, Lucy faces a data cleaning task. The microbiology lab provides data for only a subset of antibiotics based on domain characteristics and institutional preferences. At point-of-care, physicians use domain knowledge to infer susceptibilities on the unreported values. When using this data for predictive modeling, the unreported values must be filled by domain experts. To address this, we built Icarus [27] to amplify expertise in data augmentation. Icarus *guides* the domain expert by showing them high impact data subsets for edits. It allows both direct *interactions* via edits and indirect *interaction* via rules. Finally, Icarus *accelerates* task completion by leveraging the UMLS classification to suggest general rules based on the domain expert's single edit. It also allows the domain expert to preview the impact of a rule by *summarizing* the cells that will be impacted.

Due to the subjective nature of this task, multiple domain experts need to come to consensus on unreported values. To amplify the consensus process, we designed Delphi [22], which visualizes the conflicts and redundancies in domain expert rules. It provides an overview of the dataset by visually *summarizing* the antibiotics and related rules in a node-link diagram. The node sizes *guide* the expert to regions of high conflict by encoding the number of data points affected. It allows the domain experts to *interactively* edit the rule set by accepting and rejecting rules. Finally, it *accelerates* the domain experts' task completion by automatically removing redundant rules after each edit.

Once domain experts have come to a consensus, the dataset is ready for analysis. Our data scientist uses penalized logistic regression to model the resistance[219]. During this stage, Lucy provides insights on the different variables and their relations. After model creation, Lucy can analyze and validate the results on interactive analysis. For a given patient, the system should *summarize* its results by showing the probability of coverage along with confidence intervals. It should *guide* Lucy by drawing attention to any abnormal covariates, whose value significantly deviates from others in the cohort. It should allow Lucy to *interactively* select covariates and rerun the model for the specific patient. It should *accelerate* analysis by showing similar patients for who the model should also be updated.

Discussion

We have provided examples from the informatics literature to motivate the need for domain expert involvement in all steps of clinical data pipelines, from curation to analysis. While this work is based on our experiences, we have done our best to do a targeted interdisciplinary review that can serve as a guide to clinical data projects. Our work is related to previous surveys in visual analytics in healthcare[217] and interactive systems[133]. Our survey is unique in that it focuses on the clinical data pipeline and provides a taxonomy on designing systems for amplifying expertise. Specifically, expertise amplification involves: summarization, guidance, interactivity, and acceleration. Our case study illustrates how these can be applied to a clinical data pipeline.

Table 2: Applying Amplification to the Clinical Data Pipeline for Empiric Antibiotic Prediction

		Domain Expert Task	Amplification
Data Curation		Identify variables of interest, validate patients included in the cohort and make domain-specific exclusionary rules	• *Summarization*: Present distribution of variables of interest • *Guidance*: Suggesting additional variables based on the selected ones • *Interactions*: Allow expert to select and remove data points • *Acceleration:* Suggest criteria based on the domain expert's inclusion and exclusion.
Data Cleaning	Augmentation	Fill in unreported microbiology susceptibilities with rules	• *Summarization:* Preview a rule by showing distribution of the cells that will be impacted • *Guidance*: Show high impact data subsets for edits. • *Interactions:* Direct edits on interface and indirect edits via rules. • *Acceleration:* Suggest general rules based on the domain expert's single edit.
	Validation	Validate data augmentation by examining rule set and consolidating them to remove conflicts.	• *Summarization*: Visual summary of rules and their relations • *Guidance*: Node size guides user to high-conflict areas • *Interactions:* Edit rule-set by accepting and rejecting rules • *Acceleration:* Automatically remove redundant rules
Data Analysis		Understand the model and its predictions for individuals and different patient subpopulations	• *Summarization:* Show probability of coverage with confidence interval • *Guidance:* Highlight covariates of concern • *Interactions:* Allow domain expert to select covariates to include • *Acceleration:* Show similar patients for who the model should be updated.

Conclusion

Effectively engaging domain experts is crucial for the success of data-driven workflows. We provide a novel framework for developing systems that amplify domain expertise. Amplification systems should summarize data, guide domain experts' data navigation, allow domain experts to interact and update algorithms, and finally accelerate their task by learning from their interactions. This framework draws on research from multiple computer science sub-disciplines. As we move towards data-driven workflows, interdisciplinary methods are necessary for the greatest impact. Empowering stakeholders to interact with the data directly can lead to faster and more impactful insights and decision-making, which is vital for democratizing data to benefit society.

Conflicts of Interest
None declared.

Abbreviations
CDS: Clinical Decision Support
HCI: Human-Computer Interaction
EHR: Electronic Health Record
UMLS: Unified Medical Language System

References

1. F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

2. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, ACM, 2016.

3. K. Zarifis and Y. Papakonstantinou, "ViDeTTe Interactive Notebooks," in *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, p. 2, ACM, 2018.

4. S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer, "Wrangler: Interactive visual specification of data transformation scripts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3363– 3372, ACM, 2011.

5. K. Hu, D. Orghian, and C. Hidalgo, "DIVE: A Mixed-Initiative System Supporting Integrated Data Exploration Workflows," in *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, p. 5, ACM, 2018.

6. T. Kraska, "Northstar: An Interactive Data Science System," *Proceedings of the VLDB Endowment*, vol. 11, no. 12, 2018.

7. K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer, "Voyager: Exploratory analysis via faceted browsing of visualization recommendations," *IEEE Transactions on Visualization & Computer Graphics*, pp. 1–1, 2016.

8. D. Xin, S. Macke, L. Ma, J. Liu, S. Song, and A. Parameswaran, "Helix: Holistic optimization for accelerating iterative machine learning," *arXiv preprint arXiv:1812.05762*, 2018.

9. C. Olah, A. Satyanarayan, I. Johnson, S. Carter, L. Schubert, K. Ye, and A. Mordvintsev, "The building blocks of interpretability," *Distill*, vol. 3, no. 3, p. e10, 2018.

10. M. Yeomans, A. Shah, S. Mullainathan, and J. Kleinberg, "Making sense of recommendations," *Journal of Behavioral Decision Making*, 2017.

11. B. J. Dietvorst, J. P. Simmons, and C. Massey, "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them," *Management Science*, vol. 64, no. 3, pp. 1155–1170, 2016.

12. J. Zhang, Y. Wang, P. Molino, L. Li, and D. S. Ebert, "Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 364–373, 2019.

13. F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Vaughan, and H. Wallach, "Manipulating and measuring model interpretability," *arXiv preprint arXiv:1802.07810*, 2018.

14. M. Yin, J. W. Vaughan, and H. Wallach, "Understanding the effect of accuracy on trust in machine learning models," 2019.

15. D. Haas, J. Ansel, L. Gu, and A. Marcus, "Argonaut: Macrotask crowdsourcing for complex data processing," *Proceedings of the VLDB Endowment*, vol. 8, no. 12, 2015.

16. A. Parameswaran, A. D. Sarma, H. Garcia-Molina, N. Polyzotis, and J. Widom, "Human-assisted graph search: it's okay to ask questions," *Proceedings of the VLDB Endowment*, vol. 4, no. 5, pp. 267–278, 2011.

17. A. G. Parameswaran, H. Garcia-Molina, H. Park, N. Polyzotis, A. Ramesh, and J. Widom, "Crowdscreen: Algorithms for filtering data with humans," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, pp. 361–372, ACM, 2012.

18. H. Park, R. Pang, A. Parameswaran, H. Garcia-Molina, N. Polyzotis, and J. Widom, "An overview of the deco system: data model and query language; query processing and optimization," *ACM SIGMOD Record*, 2013.

19. W. S. Lasecki, J. M. Rzeszotarski, A. Marcus, and J. P. Bigham, "The effects of sequence and delay on crowd work," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1375–1378, ACM, 2015.

20. D. Retelny, S. Robaszkiewicz, A. To, W. S. Lasecki, J. Patel, N. Rahmati, T. Doshi, M. Valentine, and M. S. Bernstein, "Expert crowdsourcing with flash teams," in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 75–85, ACM, 2014.

21. I. E. Dror, J. Kukucka, S. M. Kassin, and P.A. Zapf, "When expert decision making goes wrong: Consensus, bias, the role of experts, and accuracy," *Journal of applied research in memory and cognition*, vol. 7, no. 1, pp. 162–163, 2018.

22. P. Rahman, J. Chen, C. Hebert, P. Pancholi, M. Lustberg, K. Stevenson, and A. Nandi, "Exploratory Visualizations of Rules for Validation of Expert Decisions," in *DSIA Workshop, IEEE VIS*, 2018.

23. R. A. Robbins, "Physicians generate an average 2.4 million a year per hospital," *Southwest Journal of Pulmonary and Critical Care*, vol. 18, p. 61, 2019.

24. H. Saitwal, X. Feng, M. Walji, V. Patel, and J. Zhang, "Assessing performance of an electronic health record (EHR) using cognitive task analysis," *International Journal of Medical Informatics*, vol. 79, no. 7, pp. 501–506, 2010.

25. J. Heer, "Agency plus automation: Designing artificial intelligence into interactive systems," *Proceedings of the National Academy of Sciences*, vol. 116, no. 6, pp. 1844–1850, 2019.

26. S. Krishnan, D. Haas, M. J. Franklin, and E. Wu, "Towards reliable interactive data cleaning: a user survey and recommendations.," in *HILDA@ SIGMOD*, p. 9, 2016.

27. P. Rahman, C. Hebert, and A. Nandi, "ICARUS: Minimizing Human Effort in Iterative Data Completion," *PVLDB*, vol. 11, no. 13, 2018.

28. C. J. Cai, E. Reif, N. Hegde, J. Hipp, B. Kim, D. Smilkov, M. Wattenberg, F. Viegas, G. S. Corrado, M. C. Stumpe, *et al.*, "Human-centered tools for coping with imperfect algorithms during medical decision-making," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2019.

29. P. D. Clayton, S. P. Narus, W. A. Bowes III, T. S. Madsen, A. B. Wilcox, G. Orsmond, B. Rocha, S. N. Thornton, S. Jones, C. A. Jacobsen, *et al.*, "Physician use of electronic medical records: issues and successes with direct data entry and physician productivity," in *AMIA annual symposium proceedings*, vol. 2005, p. 141, American Medical Informatics Association, 2005.

30. A. Ganapathi, and Y. Chen, "Data quality: Experiences and lessons from operationalizing big data," in *2016 IEEE International Conference on Big Data (Big Data)*, pp. 1595–1602, IEEE, 2016.

31. S. A. Collins, E. Gesner, P.L. Mar, D. M. Colburn, and R. A. Rocha, "Prioritization and refinement of clinical data elements within EHR systems," in *AMIA Annual Symposium Proceedings*, vol. 2016, p. 421, American Medical Informatics Association, 2016.

32. J. C. Blaisure and W. M. Ceusters, "Improving the fitness for purpose of common data models through realism based ontology," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 440, American Medical Informatics Association, 2017.

33. X. Pan and J. J. Cimino, "Identifying the clinical laboratory tests from unspecified other lab test data for secondary use," in *AMIA Annual Symposium Proceedings*, vol. 2015, p. 1018, American Medical Informatics Association, 2015.

34. L. Cui, Y. Huang, S. Tao, S. D. Lhatoo, and G.-Q. Zhang, "Odacci: Ontology-guided data curation for multisite clinical research data integration in the ninds center for sudep research," in *AMIA Annual Symposium Proceedings*, vol. 2016, p. 441, American Medical Informatics Association, 2016.

35. E. S. Hall, N. Connolly, D. E. Jones, and E. A. DeFranco, "Integrating public data sets for analysis of maternal airborne environmental exposures and stillbirth," in *AMIA Annual Symposium Proceedings*, vol. 2014, p. 599, American Medical Informatics Association, 2014.

36. R. Gouripeddi, J. C. Facelli, R. L. Bradshaw, D. Schultz, B. LaSalle, P. B. Warner, R. Butcher, R. Madsen, and P. Mo, "Further: An infrastructure for clinical, translational and comparative effectiveness research.," in *AMIA*, 2013.

37. X. Chen and F. Wang, "Integrative spatial data analytics for public health studies of new york state," in *AMIA*

Annual Symposium Proceedings, vol. 2016, p. 391, American Medical Informatics Association, 2016.

38. Y. Li, *Combining Heterogeneous Databases to Detect Adverse Drug Reactions*. Ph.D. thesis, Columbia University, 2015.

39. M. D. Clarkson and M. E. Whipple, "Variation in the representation of human anatomy within digital resources: Implications for data integration," in *AMIA Annual Symposium Proceedings*, vol. 2018, p. 330, American Medical Informatics Association, 2018.

40. D. C. Berrios, A. Beheshti, and S. V. Costes, "Fairness and usability for open-access omics data systems," in *AMIA Annual Symposium Proceedings*, vol. 2018, p. 232, American Medical Informatics Association, 2018.

41. F. J. Farach, O. McGettrick, C. Tirrell, C. C. Evans, A. Mesa, and L. Rozenblit, "Rexmart: An open-source tool for exploring and sharing research data without compromising data integrity.," in *AMIA*, 2014.

42. J. A. Maldonado, M. Marcos, J. T. Fernández-Breis, E. Parcero, D. Boscá, M. del Carmen Legaz-García, B. Mart´ınez-Salvador, and M. Robles, "A platform for exploration into chaining of web services for clinical data transformation and reasoning," in *AMIA Annual Symposium Proceedings*, vol. 2016, p. 854, American Medical Informatics Association, 2016.

43. Y. Zhang, B. Tang, M. Jiang, J. Wang, and H. Xu, "Domain adaptation for semantic role labeling of clinical text," *Journal of the American Medical Informatics Association*, vol. 22, no. 5, pp. 967–979, 2015.

44. L. Cui, S. Tao, and G.-Q. Zhang, "A semantic-based approach for exploring consumer health questions using UMLS," in *AMIA Annual Symposium Proceedings*, vol. 2014, p. 432, American Medical Informatics Association, 2014.

45. X. L. Dong and T. Rekatsinas, "Data integration and machine learning: a natural synergy," in *Proceedings of the 2018 International Conference on Management of Data*, pp. 1645–1650, ACM, 2018.

46. M. Stonebraker and I. F. Ilyas, "Data integration: The current status and the way forward.," *IEEE Data Eng. Bull.*, vol. 41, no. 2, pp. 3–9, 2018.

47. R. C. Fernandez and S. Madden, "Termite: A system for tunneling through heterogeneous data," *arXiv preprint arXiv:1903.05008*, 2019.

48. S. Thirumuruganathan, N. Tang, and M. Ouzzani, "Data curation with deep learning [vision]: Towards self-driving data curation," *arXiv preprint arXiv:1803.01384*, 2018.

49. R. Leaman and Z. Lu, "Taggerone: joint named entity recognition and normalization with semi-Markov models," *Bioinformatics*, vol. 32, no. 18, pp. 2839–2846, 2016.

50. R. I. Doğan, R. Leaman, and Z. Lu, "NCBI disease corpus: a resource for disease name recognition and concept normalization," *Journal of biomedical informatics*, vol. 47, pp. 1–10, 2014.

51. C.-H. Wei, H.-Y. Kao and Z. Lu, "Gnormplus: an integrative approach for tagging genes, gene families, and protein domains," *BioMed research international*, vol. 2015, 2015.

52. M. Krallinger, O. Rabal, F. Leitner, M. Vazquez, D. Salgado, Z. Lu, R. Leaman, Y. Lu, D. Ji, D. M. Lowe, *et al.*, "The chemdner corpus of chemicals and drugs and its annotation principles," *Journal of cheminformatics*, vol. 7, no. 1, p. S2, 2015.

53. C.-H. Wei, B. R. Harris, H.-Y. Kao, and Z. Lu, "tmvar: a text mining approach for extracting sequence variants in biomedical literature," *Bioinformatics*, vol. 29, no. 11, pp. 1433–1439, 2013.

54. F. Nargesian, K. Q. Pu, E. Zhu, B. G. Bashardoost, and R. J. Miller, "Optimizing organizations for navigating data lakes," *arXiv preprint arXiv:1812.07024*, 2018.

55. F. Nargesian, E. Zhu, K. Q. Pu, and R. J. Miller, "Table union search on open data," *Proceedings of the VLDB Endowment*, vol. 11, no. 7, pp. 813–825, 2018.

56. A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, p. 160035, 2016.

57. O. Bodenreider, "The unified medical language system (UMLS): integrating biomedical terminology," *Nucleic*

acids research, vol. 32, no. Suppl 1, pp. D267–D270, 2004.

58. A. Nandi, L. Jiang, and M. Mandel, "Gestural query specification," *Proceedings of the VLDB Endowment*, vol. 7, no. 4, pp. 289–300, 2013.

59. H. Zhang, V. Raj, T. Sellam, and E. Wu, "Precision interfaces for different modalities," in *Proceedings of the 2018 International Conference on Management of Data*, pp. 1777–1780, ACM, 2018.

60. Q. Zhang, H. Zhang, and E. Wu, "Mining precision interfaces from query logs," *arXiv preprint arXiv:1904.02344*, 2019.

61. K. J. Peterson, G. Jiang, S. M. Brue, and H. Liu, "Leveraging terminology services for extract-transform-load processes: a user-centered approach," in *AMIA Annual Symposium Proceedings*, vol. 2016, p. 1010, American Medical Informatics Association, 2016.

62. B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1589–1604, 2018.

63. Z. Che, S. Purushotham, R. Khemani, and Y. Liu, "Interpretable deep models for ICU outcome prediction," in *AMIA Annual Symposium Proceedings*, vol. 2016, p. 371, American Medical Informatics Association, 2016.

64. M. Adibuzzaman, P. DeLaurentis, J. Hill, and B. D. Benneyworth, "Big data in healthcare–the promises, challenges and opportunities from a research perspective: A case study with a model database," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 384, American Medical Informatics Association, 2017.

65. M. Dallachiesa, A. Ebaid, A. Eldawy, A. Elmagarmid, I. F. Ilyas, M. Ouzzani, and N. Tang, "Nadeef: a commodity data cleaning system," in *ACM SIGMOD*, 2013.

66. C. Mayfield, J. Neville, and S. Prabhakar, "Eracer: a database approach for statistical inference and data cleaning," in *ACM SIGMOD*, 2010.

67. V. V. Meduri and P. Papotti, "Towards user-aware rule discovery," in *Information Search, Integration, and Personalization*, pp. 3–17, Springer, 2017.

68. J. Wang and N. Tang, "Dependable data repairing with fixing rules," *Journal of Data and Information Quality (JDIQ)*, vol. 8, no. 3-4, p. 16, 2017.

69. J. Wang, S. Krishnan, M. J. Franklin, K. Goldberg, T. Kraska, and T. Milo, "A sample-and-clean framework for fast and accurate query processing on dirty data," in *ACM SIGMOD*, 2014.

70. S. Krishnan, J. Wang, E. Wu, M. J. Franklin, and K. Goldberg, "Activeclean: Interactive data cleaning while learning convex loss models," *arXiv preprint arXiv:1601.03797*, 2016.

71. J. Xu, D. V. Kalashnikov, and S. Mehrotra, "Query aware determinization of uncertain objects," *IEEE Transactions on Knowledge and Data Engineering*, 2015.

72. Z. Wei and S. Link, "DataProf: semantic profiling for iterative data cleansing and business ruleacquisition," in *Proceedings of the 2018 International Conference on Management of Data*, pp. 1793–1796, ACM, 2018.

73. M. Yakout, A. K. Elmagarmid, J. Neville, M. Ouzzani, and I. F. Ilyas, "Guided data repair," *VLDB*, 2011.

74. S. Thirumuruganathan, L. Berti-Equille, M. Ouzzani, J.-A. Quiane-Ruiz, and N. Tang, "Uguide: User-guided discovery of fd-detectable errors," in *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 1385–1397, ACM, 2017.

75. H. Park and J. Widom, "Crowdfill: collecting structured data from the crowd," in *ACM SIGMOD*, 2014.

76. Z. Huang and Y. He, "Auto-Detect: Data-Driven Error Detection in Tables," in *Proceedings of the 2018 International Conference on Management of Data*, pp. 1377–1392, ACM, 2018.

77. X. Chu, J. Morcos, I. F. Ilyas, M. Ouzzani, P. Papotti, N. Tang, and Y. Ye, "Katara: A data cleaning system powered by knowledge bases and crowdsourcing," in *ACM SIGMOD*, 2015.

78. S. Hao, N. Tang, G. Li, and J. Li, "Cleaning relations using knowledge bases," in *Data Engineering (ICDE), 2017 IEEE 33rd International Conference on*, pp. 933–944, IEEE, 2017.

79. F. Biessmann, D. Salinas, S. Schelter, P. Schmidt, and D. Lange, "Deep learning for missing value imputation in tables with non-numerical data," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 2017–2025, ACM, 2018.

80. T. Rekatsinas, X. Chu, I. F. Ilyas, and C. Ré, "Holoclean: Holistic data repairs with probabilistic inference," *Proceedings of the VLDB Endowment*, vol. 10, no. 11, pp. 1190–1201, 2017.

81. X. Chu, I. F. Ilyas, and P. Papotti, "Holistic data cleaning: Putting violations into context," in *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*, pp. 458–469, IEEE, 2013.

82. L. Schreibstein, R. Newton-Dame, K. H. McVeigh, S. E. Perlman, J. Singer, T. G. Harris, and C. M. Greene, "Missing data in an electronic health record-based population health surveillance system.," in *AMIA*, 2014.

83. G. Divita, Q. T. Zeng, A. V. Gundlapalli, S. Duvall, J. Nebeker, and M. H. Samore, "Sophia: a expedient UMLS concept extraction annotator," in *AMIA Annual Symposium Proceedings*, vol. 2014, p. 467, American Medical Informatics Association, 2014.

84. L. Rumeng, N. Jagannatha Abhyuday, and Y. Hong, "A hybrid neural network model for joint prediction of presence and period assertions of medical events in clinical notes," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 1149, American Medical Informatics Association, 2017.

85. A. C. Browne, M. Kayaalp, Z. A. Dodd, P. Sagan, and C. J. McDonald, "The challenges of creating a gold standard for de-identification research," in *AMIA Annual Symposium Proceedings*, vol. 2014, p. 353, American Medical Informatics Association, 2014.

86. K. H. Bowles, S. J. Ratcliffe, M. D. Naylor, J. H. Holmes, S. K. Keim, and E. J. Flores, "Nurse generated EHR data supports post-acute care referral decision making: Development and validation of a two-step algorithm," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 465, American Medical Informatics Association, 2017.

87. C. Shivade, C. Hebert, K. Regan, E. Fosler-Lussier, and A. M. Lai, "Automatic data source identification for clinical trial eligibility criteria resolution," in *AMIA Annual Symposium Proceedings*, vol. 2016, p. 1149, American Medical Informatics Association, 2016.

88. C. Norman, M. Leeflang, and A. Névéol, "Data extraction and synthesis in systematic reviews of diagnostic test accuracy: A corpus for automating and evaluating the process," in *AMIA Annual Symposium Proceedings*, vol. 2018, p. 817, American Medical Informatics Association, 2018.

89. P. Chandar, A. Yaman, J. Hoxha, Z. He, and C. Weng, "Similarity-based recommendation of new concepts to a terminology," in *AMIA Annual Symposium Proceedings*, vol. 2015, p. 386, American Medical Informatics Association, 2015.

90. R. Kavuluru and A. Rios, "Automatic assignment of non-leaf mesh terms to biomedical articles," in *AMIA Annual Symposium Proceedings*, vol. 2015, p. 697, American Medical Informatics Association, 2015.

91. D. J. Feller, J. Zucker, *et al.*, "Towards the inference of social and behavioral determinants of sexual health: Development of a gold-standard corpus with semi-supervised learning," in *AMIA Annual Symposium Proceedings*, vol. 2018, p. 422, American Medical Informatics Association, 2018.

92. M. Afshar, C. Joyce, A. Oakey, P. Formanek, P. Yang, M. M. Churpek, R. S. Cooper, S. Zelisko, R. Price, and D. Dligach, "A computable phenotype for acute respiratory distress syndrome using natural language processing and machine learning," in *AMIA Annual Symposium Proceedings*, vol. 2018, p. 157, American Medical Informatics Association, 2018.

93. A. J. Ratner, S. H. Bach, H. R. Ehrenberg, and C. Ré, "Snorkel: Fast training set generation for information extraction," in *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 1683–1686, ACM, 2017.

94. Y. Halpern, Y. Choi, S. Horng, and D. Sontag, "Using anchors to estimate clinical state without labeled data," in *AMIA Annual Symposium Proceedings*, vol. 2014, p. 606, American Medical Informatics Association, 2014.

95. V. Raman and J. Hellerstein, "Potter?s wheel: An interactive framework for data cleaning," tech. rep., Working Paper, 1999. http://www. cs. berkeley. edu/~ rshankar/papers/pwheel. pdf, 2000.

96. C. Stolte, D. Tang, and P. Hanrahan, "Polaris: A system for query, analysis, and visualization of

multidimensional relational databases," *IEEE*, 2002.

97. P. J. Guo, S. Kandel, J. M. Hellerstein, and J. Heer, "Proactive wrangling: mixed-initiative end-user programming of data transformation scripts," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pp. 65–74, ACM, 2011.

98. R. Singh and S. Gulwani, "Learning semantic string transformations from examples," *Proceedings of the VLDB Endowment*, vol. 5, no. 8, pp. 740–751, 2012.

99. Z. Jin, M. R. Anderson, M. Cafarella, and H. Jagadish, "Foofah: Transforming data by example," in *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 683–698, ACM, 2017.

100. M.-F. Costabile, D. Fogli, C. Letondal, P. Mussio, and A. Piccinno, "Domain-expert users and their needs of software development," in *HCI 2003 End-User Development Session*, 2003.

101. D. A. Hanauer, G. W. Hruby, D. G. Fort, L. V. Rasmussen, E. A. Mendonça, and C. Weng, "What is asked in clinical data request forms? A multi-site thematic analysis of forms towards better data access support," in *AMIA Annual Symposium Proceedings*, vol. 2014, p. 616, American Medical Informatics Association, 2014.

102. G. W. Hruby, J. Hoxha, P. C. Ravichandran, E. A. Mendonça, D. A. Hanauer, and C. Weng, "A data-driven concept schema for defining clinical research data needs," *International Journal of medical informatics*, vol. 91, pp. 1–9, 2016.

103. T. Edinger, D. Demner-Fushman, A. M. Cohen, S. Bedrick, and W. Hersh, "Evaluation of clinical text segmentation to facilitate cohort retrieval," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 660, American Medical Informatics Association, 2017.

104. Z. Hu, G. B. Melton, N. D. Moeller, E. G. Arsoniadis, Y. Wang, M. R. Kwaan, E. H. Jensen, and G. J. Si- mon, "Accelerating chart review using automated methods on electronic health record data for postoperative complications," in *AMIA Annual Symposium Proceedings*, vol. 2016, p. 1822, American Medical Informatics Association, 2016.

105. V. Major, M. S. Tanna, S. Jones, and Y. Aphinyanaphongs, "Reusable filtering functions for application in ICU data: a case study," in *AMIA Annual Symposium Proceedings*, vol. 2016, p. 844, American Medical Informatics Association, 2016.

106. Z. Zhao, L. De Stefani, E. Zgraggen, C. Binnig, E. Upfal, and T. Kraska, "Controlling false discoveries during interactive data exploration," in *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 527–540, ACM, 2017.

107. S. Romero-Brufau, P. Kostandy, K. L. Maass, P. Wutthisirisart, M. Sir, B. Bartholmai, M. Stuve, and K. Pasupathy, "Development of data integration and visualization tools for the department of radiology to display operational and strategic metrics," in *AMIA Annual Symposium Proceedings*, vol. 2018, p. 942, American Medical Informatics Association, 2018.

108. M. Pore, D. M. Sengeh, P. Mugambi, N. V. Purswani, T. Sesay, A. L. Arnold, A.-M. A. Tran, and R. Myers, "Design and evaluation of a web-based decision support tool for district-level disease surveillance in a low- resource setting," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 1401, American Medical Informatics Association, 2017.

109. G. Iyer, S. DuttaDuwarah, and A. Sharma, "Datascope: Interactive visual exploratory dashboards for large multidimensional data," in *2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, pp. 17–23, IEEE, 2017.

110. F. J. Anscombe, "Graphs in statistical analysis," *The American Statistician*, vol. 27, no. 1, pp. 17–21, 1973.

111. B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *The craft of information visualization*, pp. 364–371, Elsevier, 2003.

112. M. Vartak, S. Madden, A. Parameswaran, and N. Polyzotis, "SEEDB: Automatically Generating Query Visualizations," *Proceedings of the VLDB Endowment*, vol. 7, no. 13, pp. 1581–1584, 2014.

113. D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer, "Formalizing visualization design knowledge as constraints: Actionable and extensible models in Draco," *IEEE transactions on visualization and*

computer graphics, vol. 25, no. 1, pp. 438–448, 2019.

114. C Demiralp, P. J. Haas, S. Parthasarathy, and T. Pedapati, "Foresight: Rapid Data Exploration Through Guideposts," *arXiv preprint arXiv:1709.10513*, 2017.

115. H. S. G. Caballero, A. Corvo, P. M. Dixit, and M. A. Westenberg, "Visual analytics for evaluating clinical pathways," in *2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, pp. 39–46, IEEE, 2017.

116. W. Widanagamaachchi, Y. Livnat, P.-T. Bremer, S. Duvall, and V. Pascucci, "Interactive visualization and exploration of patient progression in a hospital setting," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 1773, American Medical Informatics Association, 2017.

117. X. Li, L. Cui, S. Tao, N. Zeng, and G.-Q. Zhang, "Spindlesphere: a web-based platform for large-scale sleep spindle analysis and visualization," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 1159, American Medical Informatics Association, 2017.

118. J. M. Mortensen, M. A. Musen, and N. F. Noy, "An empirically derived taxonomy of errors in snomed ct," in *AMIA Annual Symposium Proceedings*, vol. 2014, p. 899, American Medical Informatics Association, 2014.

119. E. S. Chen, G. B. Melton, R. C. Wasserman, P. T. Rosenau, D. B. Howard, and I. N. Sarkar, "Mining and visualizing family history associations in the electronic health record: A case study for pediatric asthma," in *AMIA Annual Symposium Proceedings*, vol. 2015, p. 396, American Medical Informatics Association, 2015.

120. P. S. Sockolow, Y. Yang, E. J. Bass, K. H. Bowles, A. Holmberg, and P. Sheryl, "Data visualization of home care admission nurses decision-making," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 1597, American Medical Informatics Association, 2017.

121. E. Kummerfeld, J. A. Anker, A. Rix, and M. G. Kushner, "Methodological advances in the study of hidden variables: A demonstration on clinical alcohol use disorder data," in *AMIA Annual Symposium Proceedings*, vol. 2018, p. 710, American Medical Informatics Association, 2018.

122. S. Sarawagi, R. Agrawal, and N. Megiddo, "Discovery-driven exploration of OLAP data cubes," in *International Conference on Extending Database Technology*, pp. 168–182, Springer, 1998.

123. M. Sordo, P. Tokachichu, C. J. Vitale, S. M. Maviglia, and R. A. Rocha, "Modeling contextual knowledge for clinical decision support," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 1617, American Medical Informatics Association, 2017.

124. S. Gangadhar, N. Nguyen, J. W. Pesuit, A. N. Bogdanov, L. Kallenbach, J. Ken, J. Vasey, and R. M. Loomis, "Effectiveness of a cloud-based EHR clinical decision support program for body mass index (BMI) screening and follow-up," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 742, American Medical Informatics Association, 2017.

125. S. B. Souissi, M. Abed, L. Elhiki, P. Fortemps, and M. Pirlot, "Reducing the toxicity risk in antibiotic prescriptions by combining ontologies with a multiple criteria decision model," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 1625, American Medical Informatics Association, 2017.

126. S. D. Cardoso, R.-D. Chantal, M. Da Silveira, and C. Pruski, "Combining rules, background knowledge and change patterns to maintain semantic annotations," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 505, American Medical Informatics Association, 2017.

127. M. Hedda, B. A. Malin, C. Yan, and D. Fabbri, "Evaluating the effectiveness of auditing rules for electronic health record systems," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 866, American Medical Informatics Association, 2017.

128. R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1721–1730, ACM, 2015.

129. Z. Che, J. S. Sauver, H. Liu, and Y. Liu, "Deep learning solutions for classifying patients on opioid use," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 525, American Medical Informatics Association, 2017.

130. W. Ge, J.-W. Huh, Y. R. Park, J.-H. Lee, Y.-H. Kim, and A. Turchin, "An interpretable ICU mortality prediction model based on logistic regression and recurrent neural networks with lstm units.," in *AMIA Annual Symposium Proceedings*, vol. 2018, p. 460, American Medical Informatics Association, 2018.

131. K. C. Ho, W. Speier, S. El-Saden, and C. W. Arnold, "Classifying acute ischemic stroke onset time using deep imaging features," in *AMIA Annual Symposium Proceedings*, vol. 2017, p. 892, American Medical Informatics Association, 2017.

132. Y. Ming, H. Qu, and E. Bertini, "Rulematrix: Visualizing and understanding classifiers with rules," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 342–352, 2019.

133. Rahman P, Jiang L, Nandi A. Evaluating interactive data systems. The VLDB Journal. 2020 Jan 1;29(1):119-46.

134. A. Doan, P. Domingos, and A. Y. Halevy, "Reconciling schemas of disparate data sources: A machine-learning approach," in *ACM Sigmod Record*, vol. 30, pp. 509–520, ACM, 2001.

135. H.-H. Do and E. Rahm, "Coma: a system for flexible combination of schema matching approaches," in *Proceedings of the 28th international conference on Very Large Data Bases*, pp. 610–621, VLDB Endowment, 2002.

136. K. Chen, A. Kannan, J. Madhavan, and A. Halevy, "Exploring schema repositories with schemr," *ACM SIG-MOD Record*, vol. 40, no. 1, pp. 11–16, 2011.

137. E. Rahm and P. A. Bernstein, "A survey of approaches to automatic schema matching," *the VLDB Journal*, vol. 10, no. 4, pp. 334–350, 2001.

138. A. Nandi and P. A. Bernstein, "Hamster: using search clicklogs for schema and taxonomy matching," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 181–192, 2009.

139. Y. Wang and Y. He, "Synthesizing mapping relationships using table corpus," in *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 1117–1132, ACM, 2017.

140. Z. Ives, C. Knoblock, S. Minton, M. Jacob, P. Talukdar, R. Tuchinda, J. L. Ambite, M. Muslea, and C. Gazen, "Interactive data integration through smart copy & paste," *arXiv preprint arXiv:0909.1769*, 2009.

141. J. Wang, T. Kraska, M. J. Franklin, and J. Feng, "Crowder: Crowdsourcing entity resolution," *Proceedings of the VLDB Endowment*, vol. 5, no. 11, pp. 1483–1494, 2012.

142. M. Cafarella, A. Halevy, H. Lee, J. Madhavan, C. Yu, D. Z. Wang, and E. Wu, "Ten years of webtables," *Proceedings of the VLDB Endowment*, vol. 11, no. 12, pp. 2140–2149, 2018.

143. X. L. Dong, A. Halevy, and C. Yu, "Data integration with uncertainty," *The VLDB Journal*, vol. 18, no. 2, pp. 469–500, 2009.

144. C. Zhang, L. Chen, H. Jagadish, M. Zhang, and Y. Tong, "Reducing uncertainty of schema matching via crowdsourcing with accuracy rates," *IEEE Transactions on Knowledge and Data Engineering*, 2018.

145. M. J. Cafarella, A. Halevy, and N. Khoussainova, "Data integration for the relational web," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 1090–1101, 2009.

146. J. Madhavan, S. R. Jeffery, S. Cohen, X. L. Dong, D. Ko, C. Yu, and A. Halevy, "Web-scale data integration: You can only afford to pay as you go," 2007.

147. G. Limaye, S. Sarawagi, and S. Chakrabarti, "Annotating and searching web tables using entities, types and relationships," *Proceedings of the VLDB Endowment*, vol. 3, no. 1-2, pp. 1338–1347, 2010.

148. R. J. Miller, "Open data integration," *Proceedings of the VLDB Endowment*, vol. 11, no. 12, pp. 2130–2139, 2018.

149. R. C. Fernandez, E. Mansour, A. A. Qahtan, A. Elmagarmid, I. Ilyas, S. Madden, M. Ouzzani, M. Stonebraker, and N. Tang, "Seeping semantics: Linking datasets using word embeddings for data discovery," in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pp. 989–1000, IEEE, 2018.

150. J. He, E. Veltri, D. Santoro, G. Li, G. Mecca, P. Papotti, and N. Tang, "Interactive and deterministic data cleaning," in *Proceedings of the 2016 International Conference on Management of Data*, pp. 893–907, ACM, 2016.

151. M. Bergman, T. Milo, S. Novgorodov, and W.-C. Tan, "Query-oriented data cleaning with oracles," in *ACM SIGMOD*, 2015.

152. A. Assadi, T. Milo, and S. Novgorodov, "Cleaning Data with Constraints and Experts," in *Proceedings of the 21st International Workshop on the Web and Databases*, p. 1, ACM, 2018.

153. W. Fan, F. Geerts, J. Li, and M. Xiong, "Discovering conditional functional dependencies," *IEEE*, 2011.

154. F. Chiang and R. J. Miller, "Discovering data quality rules," *VLDB*, 2008.

155. L. Golab, H. Karloff, F. Korn, D. Srivastava, and B. Yu, "On generating near-optimal tableaux for conditional functional dependencies," *VLDB*, 2008.

156. G. Cong, W. Fan, F. Geerts, X. Jia, and S. Ma, "Improving data quality: Consistency and accuracy," in *VLDB*, 2007.

157. W. Fan and F. Geerts, "Foundations of data quality management," *Synthesis Lectures on Data Management*, 2012.

158. D. Z. Wang, X. L. Dong, A. D. Sarma, M. J. Franklin, and A. Y. Halevy, "Functional dependency generation and applications in pay-as-you-go data integration systems." in *WebDB*, 2009.

159. J. Rammelaere and F. Geerts, "Explaining repaired data with CFDs," *Proceedings of the VLDB Endowment*, vol. 11, no. 11, pp. 1387–1399, 2018.

160. I. F. Ilyas, V. Markl, P. Haas, P. Brown, and A. Aboulnaga, "Cords: automatic discovery of correlations and soft functional dependencies," in *Proceedings of the 2004 ACM SIGMOD international conference on Management of Data*, 647–658, ACM, 2004.

161. N. Asghar and A. Ghenai, "Automatic discovery of functional dependencies and conditional functional dependencies: A comparative study," 2015.

162. C. De Sa, A. Ratner, C. Ré, J. Shin, F. Wang, S. Wu, and C. Zhang, "Deepdive: Declarative Knowledge Base Construction," *ACM SIGMOD Record*, vol. 45, no. 1, pp. 60–67, 2016.

163. P. Varma and C. Ré, "Snuba: automating weak supervision to label training data," *Proceedings of the VLDB Endowment*, vol. 12, no. 3, pp. 223–236, 2018.

164. C. Felix, A. Dasgupta, and E. Bertini, "The exploratory labeling assistant: Mixed-initiative label curation with large document collections," in *The 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 153–164, ACM, 2018.

165. C.-H. Wei, H.-Y. Kao, and Z. Lu, "Pubtator: a web-based text mining tool for assisting biocuration," *Nucleic acids research*, vol. 41, no. W1, pp. W518–W522, 2013.

166. D. Kwon, S. Kim, C.-H. Wei, R. Leaman, and Z. Lu, "eztag: tagging biomedical concepts via interactive learning," *Nucleic acids research*, vol. 46, no. W1, pp. W523–W529, 2018.

167. A. Nazi, B. Ding, V. Narasayya, and S. Chaudhuri, "Efficient estimation of inclusion coefficient using hyperloglog sketches," *Proceedings of the VLDB Endowment*, vol. 11, no. 10, pp. 1097–1109, 2018.

168. X. Yuan, X. Cai, M. Yu, C. Wang, Y. Zhang, and Y. Wen, "Efficient foreign key discovery based on nearest neighbor search," in *International Conference on Web-Age Information Management*, pp. 443–447, Springer, 2015.

169. J. Motl and P. Kord'ık, "Foreign key constraint identification in relational databases," 2017.

170. H. Köhler and S. Link, "Inclusion dependencies reloaded," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1361–1370, ACM, 2015.

171. Z. Chen, V. Narasayya, and S. Chaudhuri, "Fast foreign-key detection in Microsoft SQL Server PowerPivot for Excel," *Proceedings of the VLDB Endowment*, vol. 7, no. 13, pp. 1417–1428, 2014.

172. D. Moritz, B. Howe, and J. Heer, "Falcon: Balancing interactive latency and resolution sensitivity for scalable linked visualizations," 2019.

173. N. Kamat, P. Jayachandran, K. Tunga, and A. Nandi, "Distributed and interactive cube exploration," in *2014 IEEE 30th International Conference on Data Engineering*, pp. 472–483, IEEE, 2014.

174. L. Lins, J. T. Klosowski, and C. Scheidegger, "Nanocubes for real-time exploration of spatiotemporal datasets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2456–2465, 2013.

175. C. A. Pahins, S. A. Stephens, C. Scheidegger, and J. L. Comba, "Hashedcubes: Simple, low memory, real-time visual exploration of big data," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 671–680, 2017.

176. M. Joglekar, H. Garcia-Molina, and A. Parameswaran, "Interactive data exploration with smart drill-down," in *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pp. 906–917, IEEE, 2016.

177. K. Dimitriadou, O. Papaemmanouil, and Y. Diao, "Aide: an active learning-based approach for interactive data exploration," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 11, pp. 2842–2856, 2016.

178. F. Psallidas and E. Wu, "Smoke: Fine-grained lineage at interactive speed," *Proceedings of the VLDB Endowment*, vol. 11, no. 6, pp. 719–732, 2018.

179. E. Wu and S. Madden, "Scorpion: Explaining away outliers in aggregate queries," *Proceedings of the VLDB Endowment*, vol. 6, no. 8, pp. 553–564, 2013.

180. S. Kandel, R. Parikh, A. Paepcke, J. M. Hellerstein, and J. Heer, "Profiler: Integrated statistical analysis and visualization for data quality assessment," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 547–554, ACM, 2012.

181. M. Correll, M. Li, G. Kindlmann, and C. Scheidegger, "Looks good to me: Visualizations as sanity checks," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 830–839, 2019.

182. K. Wongsuphasawat, Z. Qu, D. Moritz, R. Chang, F. Ouk, A. Anand, J. Mackinlay, B. Howe, and J. Heer, "Voyager 2: Augmenting visual analysis with partial view specifications," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 2648–2659, ACM, 2017.

183. W. Willett, J. Heer, and M. Agrawala, "Scented widgets: Improving Navigation Cues with Embedded Visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1129–1136, 2007.

184. A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable ai systems for the medical domain?" *arXiv preprint arXiv:1712.09923*, 2017.

185. G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, vol. 73, pp. 1–15, 2018.

186. S. Alspaugh, N. Zokaei, A. Liu, C. Jin, and M. A. Hearst, "Futzing and moseying: Interviews with professional data analysts on exploration practices," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 22–31, 2018.

187. J. Stoyanovich, B. Howe, H. Jagadish, and G. Miklau, "Panel: a debate on data and algorithmic ethics," *Proceedings of the VLDB Endowment*, vol. 11, no. 12, pp. 2165–2167, 2018.

188. C. J. Cai, J. Jongejan, and J. Holbrook, "The effects of example-based explanations in a machine learning interface," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 258–262, ACM, 2019.

189. D. Liu, P. Xu, and L. Ren, "Tpflow: Progressive partition and multidimensional pattern extraction for large-scale Spatio-temporal data analysis," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 1–11, 2019.

190. M. Khan, L. Xu, A. Nandi, and J. M. Hellerstein, "Data tweening: incremental visualization of data transforms," *Proceedings of the VLDB Endowment*, vol. 10, no. 6, pp. 661–672, 2017.

191. N. Shah, M. E. Acree, C. Patros, M. Suseno, J. Grant, G. Fleming, B. Hadsell, J. Semel, C. Hebert, K. Singh, *et al.*, "188. a novel inpatient antibiotic stewardship assistance program (asap) using real-time electronic health record data, prediction modeling, and epidemiologic data to provide personalized empiric antibiotic recommendations," in *Open Forum Infectious Diseases*, vol. 5, pp. S83–S83, Oxford University Press US, 2018.

192. N. Kamat and A. Nandi, "A session-based approach to fast-but-approximate interactive data cube exploration," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 12, no. 1, p. 9, 2018.

193. L. Battle, R. Chang, and M. Stonebraker, "Dynamic prefetching of data tiles for interactive visualization," in *Proceedings of the 2016 International Conference on Management of Data*, pp. 1363–1375, ACM, 2016.

194. L. Takayama and E. Kandogan, "Trust as an underlying factor of system administrator interface choice," in *CHI'06 extended abstracts on Human factors in computing systems*, pp. 1391–1396, ACM, 2006.

195. S. Idreos and E. Liarou, "dbtouch: Analytics at your fingertips.," in *CIDR*, 2013.

196. M. Bendre, B. Sun, D. Zhang, X. Zhou, K. C.-C. Chang, and A. Parameswaran, "DataSpread: Unifying Databases and Spreadsheets," *VLDB*, 2015.

197. E. Bakke and D. R. Karger, "Expressive query construction through direct manipulation of nested relational results," in *Proceedings of the 2016 International Conference on Management of Data*, pp. 1377–1392, ACM, 2016.

198. J. Schaffer, C. Playa Vista, J. ODonovan, J. Michaelis, A. Raglin, and T. Höllerer, "I can do better than your ai: expertise and explanations," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 240–251, ACM, 2019.

199. V. Arnold, N. Clark, P. A. Collier, S. A. Leech, and S. G. Sutton, "The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions," *Mis Quarterly*, pp. 79–97, 2006.

200. I. E. Dror, "A hierarchy of expert performance," *Journal of Applied Research in Memory and Cognition*, vol. 5, no. 2, pp. 121–127, 2016.

201. W. Mason and S. Suri, "Conducting behavioral research on amazon's Mechanical Turk," *Behavior research methods*, vol. 44, no. 1, pp. 1–23, 2012.

202. A. Kittur, E. H. Chi, and B. Suh, "Crowdsourcing user studies with Mechanical Turk," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 453–456, ACM, 2008.

203. A. Das Sarma, A. Parameswaran, and J. Widom, "Towards globally optimal crowdsourcing quality management: The uniform worker setting," in *Proceedings of the 2016 International Conference on Management of Data*, pp. 47–62, ACM, 2016.

204. E. Kandogan, M. Roth, P.Schwarz, J. Hui, I. Terrizzano, C. Christodoulakis, and R. J. Miller, "Labbook: Metadata-driven social collaborative data analysis," in *2015 IEEE International Conference on Big Data (Big Data)*, pp. 431–440, IEEE, 2015.

205. J. M. Hellerstein, V. Sreekanti, J. E. Gonzalez, J. Dalton, A. Dey, S. Nag, K. Ramachandran, S. Arora, A. Bhattacharyya, S. Das, *et al.*, "Ground: A data context service.," in *CIDR*, 2017.

206. S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer, "Enterprise data analysis and visualization: An interview study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2917–2926, 2012.

207. E. Kandogan, A. Balakrishnan, E. M. Haber, and J. S. Pierce, "From data to insight: work practices of analysts in the enterprise," *IEEE computer graphics and applications*, vol. 34, no. 5, pp. 42–50, 2014.

208. H. Jagadish *et al.*, "Organic databases," *Databases in Networked Information Systems*, 2011.

209. P. J. Embi, T. R. Yackel, J. R. Logan, J. L. Bowen, T. G. Cooney, and P. N. Gorman, "Impacts of computerized physician documentation in a teaching hospital: perceptions of faculty and resident physicians," *Journal of the American Medical Informatics Association*, vol. 11, no. 4, pp. 300–309, 2004.

210. P. Rahman and A. Nandi, "Transformer: a database-driven approach to generating forms for constrained interaction," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 485–496, ACM, 2019.

211. K. Chen, H. Chen, N. Conway, J. M. Hellerstein, and T. S. Parikh, "Usher: Improving data quality with dynamic forms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 8, pp. 1138–1153, 2011.

212. K. Gajos and D. S. Weld, "Supple: automatically generating user interfaces," in *Proceedings of the 9th international conference on Intelligent user interfaces*, pp. 93–100, ACM, 2004.

213. M. Jayapandian and H. Jagadish, "Automated generation of forms-based database query interface," *VLDB*, 2008.

214. Hielscher T, Niemann U, Preim B, Völzke H, Ittermann T, Spiliopoulou M. A framework for expert-driven subpopulation discovery and evaluation using subspace clustering for epidemiological data. Expert Systems with Applications. 2018 Dec 15;113:147-60.

215. Zhang Z, Gotz D, Perer A. Iterative cohort analysis and exploration. Information Visualization. 2015 Oct;14(4):289-307.

216. Malik S, Du F, Monroe M, Onukwugha E, Plaisant C, Shneiderman B. An evaluation of visual analytics approaches to comparing cohorts of event sequences. InEHRVis Workshop on Visualizing Electronic Health Record Data at VIS 2014 Nov 9 (Vol. 14).

217. Preim B, Lawonn K. A survey of visual analytics for public health. In Computer Graphics Forum 2020 Feb (Vol. 39, No. 1, pp. 543-580).

218. Dzau VJ, Kirch DG, Nasca TJ. To care is human—collectively confronting the clinician-burnout crisis. N Engl J Med. 2018 Jan 25;378(4):312-4.

219. Hebert C, Gao Y, Rahman P, Dewart C, Lustberg M, Pancholi P, Stevenson K, Shah N, Hade EM. Prediction of antibiotic susceptibility for urinary tract infection in a hospital setting. Antimicrobial Agents and Chemotherapy. 2020 Apr 20.