Gina Stalica
Independent Study Reflection
Digital and Computational Studies
Spring 2016

Introduction

Databases allow for the organization of information. As our modern world continues to produce more and more information, databases have become extremely useful tools; however, they should not be a replacement for other methods of storing and understanding knowledge. Like other computational tools, databases can supplement academic or other communities who share information, but they should not be a sole method of maintaining knowledge, as the rigidity and inherent assumptions they require inherently result in a loss of some degree of meaning. In order to preserve the full meaning of the objects and ideas to be studied in any discipline or community, more subjective, fluid methods of storing information are necessary. Until such methods are developed, the use of databases as a computational tool will require the work of humanists who will maintain, understand, and share knowledge alongside the organization of information via databases and other computational tools.

This semester, I have had the opportunity to explore databases with both technological and sociological approaches. Through the lens of a Computer Science student, I have explored the processes of creating databases and connecting to servers. To fully understand how databases function, it was necessary to dive into the syntax of Structured Query Languages and the process of working with a server. Furthermore, I have researched and critically analyzed the ways in which databases are currently being used, allowing me to understand their power and their limitations.

Ultimately, my work has confirmed my belief that databases are necessary tools for use in our increasingly data-driven world, but that they should be used with great caution. The classification schemes required by databases are not new, nor are they neutral. Instead, classification schemes, as they have been used for all of humanity, require abstractions and simplifications that reproduce biases. A best practice for the use of databases calls for thorough, thoughtful consideration.

Methods

For my project, I began with a review of various MySQL and database tutorials available online. My goal was to select resources that would be useful to myself and to students at all levels of programming experience so as to be able to provide guidance to appropriate tutorials. After exploring a variety of resources, a selected a few that captured an introduction to MySQL at a variety of levels. I wrote brief reviews of each resource that includes a "Pros and Cons" list for each, in an effort to provide easily accessible direction to resources. These reviews now exist on the Bowdoin Digital and Computational Studies blog so that they may be widely available to Bowdoin students.

After reviewing tutorial resources, I moved on to create my own tutorials. Through discussion with Crystal Hall and Stephen Houser, we concluded that MySQL was the most relevant Structured Query Language to explore and that the Sequel Pro database management application was the most appropriate tool with which to explore MySQL. I created step-by-step walkthroughs, assuming Bowdoin Digital and Computational Studies students to be my primary

audience. Because Digital and Computational Studies classes provide students with access to classroom Apple MacBooks, we also decided to make the tutorials specific to Mac computers.

The first tutorial I created guides users in downloading and setting up MySQL and a MySQL server instance on their personal machine. My original tutorial is complemented by the second tutorial I created, which walks users through the process of connecting to a MySQL server and creating a new database with user credentials.

I selected my tutorial content based on my intention to prepare students to be able to create and understand their own databases. Rather than focusing strictly on the syntax of Structured Query Languages, I wanted to provide step-by-step resources that attempt to make users more comfortable with the idea of databases by working to demystify their creation and accessibility. By guiding users in creating their own personal server instances and practice tables, I aim to make databases and their design seem more approachable.

Content

Technical artifacts possess politics that result from the social systems in which they are embedded (Winner, 1980, p. 122). Databases certainly possess politics based on the context in which they are created and used. At their core, databases are structures for organizing and storing information. They are composed of tables which store entities, as represented by rows, and attributes, as represented by columns. In order to design a database, creators must consciously decide how to organize these elements. In order to understand the inherent biases that are reproduced within databases, it is important to review classification schemas.

Classification itself is an element of human nature (Bowker & Star, 1999, p.1). Systems of classification structure the world around us, in ways we may or may not recognize (Bowker & Star, 1999, p. 1). Because classification has and continues to be a natural part of human life, it can be difficult to recognize the extent of systems of classification that currently exist. Further, it is simple to look past the ways in which we classify the things we encounter in the world around us.

Bowker and Star define two types of classification: Aristotelian classification and prototype classification (1999, p. 62). Aristotelian classification systems use binary traits to determine whether or not an object or idea fits into a category (Bowker & Star, 1999, p. 62). For example, we might imagine the following Aristotelian classification system: a Ford pickup truck is identified first as an object, then as a vehicle, then as a truck, and most specifically, a Ford.

Prototype classifications are based on the use of a general imagination of an object that can be used for comparison with an object to be classified (Bowker & Star, 1999, p. 62). To use a prototype classification to identify a Ford pickup truck, one might simply have a mental image of a motor vehicle with a large bed in the rear. When an object that matches the truck is encountered in the real world, a classification occurs. Rather than progressing through a system of binary characteristics in order to come to a most specific classification, prototype classification works through a more general process.

Though the process of classification via an Aristotelian system may appear to be objective, binary systems are inherently limited, allowing for potentially harmful abstractions of information. Even binary systems may be created with the bias of those who create the classification schemas to be used or, even more broadly, the bias of the social structures under

which the schemas are produced (Bowker, & Star, 1999, p. 53). The work of David Bloor supports Émile Durkheim and Marcel Mauss' original hypothesis, that the categories created by Robert Boyle and Isaac Newton were shaped by and reflective of their own personal religious and political beliefs (Bowker & Star, 1999, p.60). This case is particularly interesting, as these Enlightenment figures have previously been believed to be the forefathers of objectivity. If even those reputed for their scientific neutrality are capable of inserting their personal agendas into the categories they created, it is clear that the biases in any classification schema should be examined (Bowker & Star, 1999, p. 60).

In order to evaluate the ability to store and share knowledge in databases, it is helpful to distinguish between knowledge and information using the data-information-knowledge-wisdom (DIKW) model, as studied by Rowley (2007, p. 163). According to the DIKW model, data exists in the form of numbers, objects, or measurements and is the result of observation (Rowley, 2007, p. 166). Information is deduced using data to establish relationships and answer questions (Rowley, 2007, p. 166). Knowledge uses information to create meaningful relationships, drawing upon personal experience of the shared experience of another in order to do so (Rowley, 2007, p. 166). Using these definitions, it can be confirmed that databases store information, but cannot hold the meaning of knowledge. According to Christie (2004), "Databases do not contain knowledge, they contain information (ie ones and zeros in particular formation). Education is not the transmission of information from one head to another (see (Reddy 1979)), it is the negotiated production of knowledge in context (Turnbull 1997)." (p.1). According to this definition, education is the creation of knowledge and the sharing of knowledge cannot depend exclusively on the use of a database.

The limitations of knowledge storage and sharing becomes extremely apparent in the study of aboriginal knowledge, as conducted by Michael Christie (2004). Members of many Aboriginal communities hold a growing concern that their younger generations lack a deeper understanding of their indigenous identity, especially as supported through indigenous knowledge (Christie, 2004, p. 1). These concerned individuals would like to be able to use databases and other digital technologies to facilitate the sharing of knowledge with younger generations; however, as Christie (2004) states, the inability for databases to store actual knowledge is somewhat limiting (p.1). The database, Christie (2004) argues, is only an artefact, serving as a placeholder for a previous assertion of true knowledge (p.1). For example, the Yolngu Aboriginal people Christie (2004) studied place utmost importance on one's identity, which they believe depends on one's personal telling of their own story (p. 2). An identity, for the Yolngu, cannot be represented in a database; it must be personally shared (Christie, 2004, p.2). The example of Yolngu identity provides a clear distinction between knowledge and information: any attempt to store or share one's identity in a database would be lacking its true meaning as held in the personal experience of being oneself.

Christie's (2004) study of aboriginal knowledge-sharing is just one example of the way in which databases can be an extremely powerful tool in storing information, but can only be truly useful in sharing knowledge if used alongside other teaching media, just as a book would not be used as an educational tool entirely on its own (p. 1). Additionally, Christie (2004) calls for databases to be read in company with the information it holds, including who owns the information, the intended use of the information, and more (p. 3). Ideally, databases should be

designed in collaboration with community for which it will serve and be used, changing in its structure as deemed necessary (Christie, 2004, p.3).

In terms of content, databases may be criticized for their inherent biases and their limitations. While the nature of databases may be unavoidably limited in their ability to share knowledge, they can be supplemented in a way that can account for any meaning they may lack. Database designers should be held responsible for remaining straightforward about the assumptions and simplifications that are embedded within their tools. I believe that the greatest content issue lies in the biases reproduced by the classification schemes selected for use. This issue depends on human decision-making, bringing the importance of database users and creators to light.

Users

While databases allow for data to be stored and organized so that it can be maintained and used over time, their potential user groups are necessarily limited. Interacting with databases requires a certain degree of sociotechnical savvy. Extracting and analyzing the data held in a database requires an understanding of the databases' internal structures. For example, in order to extract a specific entry in a database, the user must know how to locate the entry's unique identifier, known as its primary key. To extract an entire attribute, or column, of data, the user must be aware of the categories that exist in the database. Designing and creating databases requires even more concrete technical skill, including an understanding of how to connected to a server and how to use a Structured Query Language, such as MySQL. In order to interact with a database at all, the user must have access to it, whether through a website or access to a server.

Independent of technical skills, database users must possess a certain degree of understanding regarding the limitations of databases in order to use them effectively. Uninformed viewers may accept the information stored in a database at "face value" (Drucker, 2012, p. 89). The creators of a given database may recognize the simplifications involved with the use of the database as a tool, but others who interact with the database may not. This may naturally result in a loss of a certain degree of meaning. For example, the International Classification of Disease (ICD), a coding system used by medical professionals, includes only a limited range of potential causes of death that can be attributed to a patient (Bowker & Star, 1999, p. 172). Medical professionals can ultimately only assign a single primary cause of death to a patient, despite the fact that a patient may have had a more complex health history that requires further context or explanation (Bowker & Star, 1999, p. 172). For example, tuberculosis is a disease whose causes and manifestations are changed constantly over time, making it difficult for professionals to diagnose or assign a cause to instances of the disease, forcing generalizations and simplifications (Bowker & Star, 1999, p. 171). If an outsider viewed a particular individual's diagnosis or cause of death, they may not understand the complete situation of an individual's health.

Even when we, as users, approach databases with a belief that they are "formal," with inherent flaws, we end up altering our behavior according to their perceived formalities (Bowker & Star, 1999, p. 53). When a limited number of classification options are available to a user, forced, improper categorization naturally occurs. In the case of the ICD, medical professionals have been and continue to be forced to make generalizations about the true conditions of a patient in order to fit the predetermined system of classification (Bowker & Star, 1999, p. 172).

In order to use databases most effectively, users must be aware of the limitations of the given system.

Ideally, users should be involved in the creation of database classification schemes. While database creators may hold the greatest understanding of the technicalities of a database, users are the true content specialists. Though traditional creators, likely information technology specialists, may find the involvement of users in database creation to be unnecessarily complicated, users can offer deeper insight as to how to best organize a database in order for it to appropriately organized its intended content. For example, in Michael Christie's study of databases and Aboriginal knowledge-sharing, Christie calls for a requirement that Aboriginal community members must be a part of the design process of any databases that will successfully be used as a tool for sharing their knowledge and identity (2004, p.8). If an outside information technology specialist attempted to create a database for the Yolngu people without any understanding of their community, values, or history, they would inherently create a tool that did not sufficiently meet the needs of the people. When users can be a part of the database design conversation, problematic simplifications and inherent biases in the classification scheme can be better avoided.

So long as they are considered critically, databases can allow for the production of primary sources that are available to anyone (Crymble, 2015, p.1). Especially by integrating databases into websites, organized data can be provided to an infinite range of users. This has the potential to make information more accessible, bridging the gap between those who do and do not have access to data. In the realm of academia, for example, databases have the power to

provide free access to archival information. By making information widely and freely available, allows for greater possibilities in the use or analysis of a set of archives.

A best practice for database design requires that users consider their classification structures critically. Ideally, it also requires input from its users. Unfortunately, the use of databases requires both skill and access. If databases can be made more widely available, as made possible through the integration of databases into websites, the limited potential user group can be expanded.

Implications

Throughout the semester, I have studied databases from a theoretical perspective. In researching critical material, it has become clear that classification systems have existed for all of humanity and have proved to be problematic throughout their existence and use. As databases have become more widely used, their implications have become more extensive and more deeply embedded into our lives.

As life becomes increasingly digitized, the implications of databases and their classification schemas are becoming particularly expansive. The implementation of unique Social Security Numbers enabled the government to maintain data unique to every single person, enabling them with the power to track the information of any United States citizen (Solove, 2004, p. 15). This was the critical step in the rise of databases as information-storing powerhouses, as databases require that each data entity possesses its own distinct identification value, known as a primary key. While there are certainly benefits to the storage of information, it is alarming that so much of our personal data is tracked and owned by corporations.

Databases can be used to define class systems, reinforcing systems of socioeconomic inequality. In Fourcade and Healy's (2013) assessment of classification situations in the neoliberal age, it is noted that social class systems have been defined by production-based categorization systems that have shaped social structures and individual experiences for as long as economic systems have existed (p. 559). Currently, credit systems use scoring technologies to place individuals into what Fourcade and Healy (2013) define as classification situations or, "positions in the credit market that are consequential for one's life-chances, and that are associated with distinctive experiences of debt." (p. 560). Because consumer behaviors can be so easily tracked, individuals can be surveilled and categorized according to their financial decisions and situations (Foucade & Healy, 2013, p. 567). When individuals are placed into "risky" credit categories, their financial options are limited, forcing them to turn to riskier banking options that have shown to increase one's risk of bankruptcy (Foucade & Healy, 2013, p. 567). Placement in "risky" categories thus becomes a way in which inequality is reproduced and perpetuated on a systematic level (Foucade & Healy, 2013, p. 567). Simultaneously, this becomes problematic on an individual level, especially when considering the adverse personal effects that are associated with this perpetuated lack of financial access, including higher rates of mental health disorders (Foucade & Healy, 2013, p. 567).

Categorization is becoming even more embedded in our lives as online data, whether in clicks, Facebook likes, or personal information, has become completely commodified. Companies who own personal online data take pride in their categorization techniques (Solove, 2004, p. 22). Focus USA maintains mailing lists for specifically labelled groups such as, "Grown But Still At Home," "Big-Spending Parents," and "Affluent Hispanics" (Solove, 2004, p.

22). The biases held in this classification scheme are explicit. Technological information-gathering has the potential to exploit personal characteristics and perpetuate harmful stereotypes. On a grander scale, they serve to strengthen unequal economic, gender, and racial systems, especially as they are used without any awareness of the people whose data is being classified.

Databases can serve to perpetuate power structures of inequality, even beyond socioeconomic status. More specifically, attempts to improve safety and security are often wrongly used to justify the reproduction of unjust racial systems. Cote-Boucher has explored the implementation of a "smart border" that claims to improve security along the United States-Canada border in an attempt to combat terrorism (2008, p. 142). The "smart border," she argues, is really a "diffuse border," whose powers of control extend well beyond the physical North American border itself (Cote-Boucher, 2008, p. 142). Through the use of technologies of control, the smart border uses racialized categorization of individuals and groups in order to assess perceived danger (Cote-Boucher, 2008, p. 142). Because the information collected at the border can be stored in databases, their information can be used and available virtually indefinitely.

When security agencies look to assess a person's level of dangerousness, they refer to the extensive information held in these databases - information that is often categorized according to religion, physical description, race, and citizenship status (Cote-Boucher, 2008, p. 150). The maintenance of information according to these categories and their use in deeming what is or is not a threat to national security perpetuates racist stereotypes. Ultimately, this information does not successfully serve its purpose of combatting terrorism. The present threats continue to exist

beyond the grasp of data. On the basis of human decision-making, people who fall certain categories are more likely to be more closely surveilled than others.

With modern technology, both online and physical data can be tracked almost infinitely. When considering the privacy of my own data, I feel most strongly about the privacy of my online data. The thought of being categorized according to my online preferences concerns me. To be labelled as a stereotype based on this data is disturbing. Rarely do I fear for the privacy of my data; however, I realize that this is not independent of the fact that I am a healthy, white, female with full American citizenship whose religious beliefs will never be deemed as a dangerous. While I may inevitable be stored in database categories that displease me, my categorization likely will not place major limits on my life, though it may essentially limit my freedom of expression. If my curious Web searches or online academic research contains content that may be flagged, my online data may categorize me in ways that hold unforeseen power over my everyday life. Perhaps the research I conducted regarding Tor and the Deep Web has caused me to be highlighted as a potential user of the online black market. Even for those who are in privileged positions, categorization systems may have dangerous implications. It is crucial to look beyond the benefits of databases as a tool and to critically analyze these systems.

Conclusions

Ultimately, databases are a useful, if not necessary, tool for maintaining and analyzing information, but they must be used with caution. Databases cannot be the sole tool for storing information, as it cannot truly contain knowledge. Instead, databases must be used as a supplement with long-established, successful methods of knowledge-sharing, such as the

maintenance of physical archives, traditional education practices, and the practices of storytelling.

For various disciplines and communities, databases can provide an invaluable way to store information to be held indefinitely. By integrating databases into websites, information can be made widely accessible, which may allow the gap between those with access and those without access to academic information to diminish. Just as knowledge cannot be transmitted using databases alone, archival information held in databases must be accompanied by additional resources.

The tutorials I have created are one small step towards integrating database understanding and design into the Bowdoin curriculum, whether as part of the Digital and Computational Studies Initiative, the Computer Science Department, or elsewhere. If the power to design and implement databases can become more easily accessible, it will be possible to work towards the production of more collaborative, appropriate databases that can potentially serve to store information with less bias.

Because databases and their organization naturally involve abstractions and assumptions, critical thinkers must be involved in their creation and maintenance in order to produce databases that can serve as tools that do not perpetuate systems of inequality. Ideally, those who hold any stake in the data held within a database should be a part of the design process. If database design can be approached with a critical mindset, databases will be invaluable to the organization of data, whether for the sharing of knowledge, the documentation of archives, or beyond, especially if more dynamic methods of database design are considered and developed.

Bibliography

Christie, M. (2004). Computer databases and Aboriginal knowledge. Retrieved from http://www.cdu.edu.au/centres/ik/pdf/CompDatAbKnow.pdf

Cote-Boucher, K. (2008). The Diffuse Border: Intelligence-Sharing, Control and Confinement along Canada's Smart Border. *Surveillance & Society, 5.* Retrieved from http://www.surveillance-and-society.org/articles5(2)/canada.pdf

Crymble, A. (2015). Does your historical collection need a database-driven website?. *Digital Humanities Quarterly*, *9*(1).

Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out : Classification and Its Consequences*. Cambridge, Mass: The MIT Press.

Fourcade, M., & Healy, K. (2013). Classification situations: Life-chances in the neoliberal era. *Accounting, Organizations and Society, 38*, 559-572. Retrieved from https://kieranhealy.org/files/papers/classification-situations.pdf

Rowley, J. E. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*. Retrieved from http://jis.sagepub.com/content/33/2/163.abstract

Solove, D. J. (2004). *The Digital Person: Technology and Privacy in the Digital Age*. New York, NY: New York University Press.

Winner, L. (1980). Do Artifacts Have Politics?. *Daedalus*, *109*(1), 121–136. Retrieved from http://www.jstor.org/stable/20024652