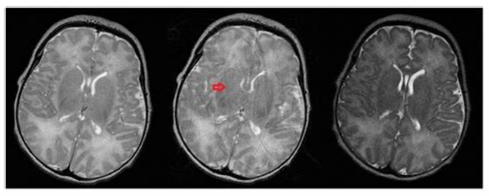
Actionable information in vision

Continue



Why visual information is important. What should be included in vision statement.

Data, information, and knowledge are terms commonly used in earth and environmental sciences, as well as in informatics supporting these sciences. The Lindstrom et al. Framework for Ocean Observing highlights the "challenge of delivering ocean information for societal benefit" and suggests that a key framework concept is to promote the "transformation of observational data organized in [Essential Ocean Variables] into information." A flyer presenting the Integrated Carbon Observation System says "Knowledge through observations." Writing about Oceans 2.0, Ocean Networks Canada highlights that the system is able to mine "data streams to detect trends, classify content and extract features [...] thereby turning raw data into information and setting the stage to allow the information to be transformed into knowledge." At 2016 AGU Fall Meeting, Rebecca Moore presented the vision of monitoring a changing planet and "generating precise, actionable information and knowledge." Yet, what exactly are these entities in the context of earth sciences and environmental research infrastructures? Can they be defined? To which processes are they input and output? How are they represented and managed? Can we extend Moore's vision to machine actionable information and knowledge? Information Systems research has for long struggled with defining data, information, and knowledge. Literature on the Data, Information, Knowledge, Wisdom (DIKW) hierarchy underscores the challenge of defining these terms. Some scholars have even suggested that providing general definitions is beyond the scope of the discipline. This may be particularly true at the higher levels, where wisdom should be considered in the context of the societal environment and may not be quantifiable out of context. While reaching consensus is hard, to obtain a better understanding for what the terms mean, how they are applied, and to what processes they are relevant in the context of earth sciences and environmental research infrastructures is arguably worthwhile. This can be done in some situations through the examination of exemplars or use cases, particularly addressing processing for translation of data to knowledge. In this talk, we will not attempt to define what data, information, and knowledge are in the context of earth sciences and environmental research infrastructures. Rather, in the particular context of a concrete use case in aerosol science - namely for the study of atmospheric new particle formation events on concentration of polydisperse aerosol - we present how observational data on concentration evolve to but are different from information about events, and how these entities are input and output, respectively, to the process of interpretation. The presentation involves technologies that enable the formal representation and management of information. Information about new particle formation events is thus machine actionable. The Situation In todays information-driven economy nowhere is clean, connected and trustworthy information more vital than in banking where information is the lifeblood of the business. In banking accurate, timely and actionable information is the difference between market leaders and also-ran. The banking business model is based on the concept of leverage. Banks raise capital through deposits, borrowings and sale of financial instruments and turn around and invest that capital to make loans and mortgages. Bank earnings are driven by the interest spread between interest earned on loans and interest paid out on deposits and borrowings. This is an overly simplistic generalization but is the crux of how the banking business model works. The Opportunity The wrinkle in the banking model is the need for reserve capital. Turns out that for everyone to have trust and confidence in the system and for the model to work, banks should maintain reserve capital to meet financial obligations incase the depositors or borrowers come looking for their money. The actual level of reserve capital is determined by government regulations but for current purposes its fair to assume that for every dollar banks raise they are required to keep up to 10 cents in reserves. Imagine what happens if the banks have dirty data like multiple instances of a single depositor or the single liability – the reserve capital requirements go up at the expense of deployed capital. Banks get hit with idle interest bearing deposits or borrowings and lower deployed capital resulting in increased cost, reduced revenues and lower margins. So, how do banks address these data quality issues? Simple – have IT use their classic, rule-based data quality tools to clean up the data, remove duplicates, standardize and everyone's happy. Not so fast! The case for business analyst-friendly, smart, enterprise-grade data preparation Look at the sample of depositor information on the right. The classic, rule-base data quality approach using traditional tools to clean, deduplicate and standardize data won't work in this case since even after cleaning customer names different addresses or phone numbers might cause these records to be flagged as unique resulting in duplicate data. The massive volume, variety and velocity of human and machine generated data and the ever changing nature of data quality issues further complicate things. The classic, rule-based data quality systems simply can't keep pace with the data deluge and unknowns. Enter business analyst-friendly, smart, self-service data preparation platform like Paxata. Here's how a modern data prep solution like Paxata addresses these challenges: Business analyst-friendly, self-service solution: Paxata provides an Excel like interface for non technical business users to interactively and visually clean, combine and transform data without writing code, sampling data or building schemas. A banking analyst with the right business context and interactively working with the depositor information can quickly spot the issues and in a few clicks automatically clean, deduplicate and standardize data. Paxata balances business's need for information with IT's need for governance so business is empowered with self-service information creation and consumption within the guardrails of governance provided by IT. Smart: With the massive volume, variety and velocity of data confronting the banks, business analyst-friendliness isn't enough. The analysts need an intelligent solution that can automatically spot and address known and unknown data quality issues at scale across multiple business units, geographic locations, customer segments and product categories. Enter Paxata's smart data preparation solution that leverages machine learning, natural language processing (NLP), semantic analysis, in-memory processing and commodity hardware based distributed processing technologies such as Spark and Hadoop to clean, combine, shape, enrich and transform data and address known and unknown data quality issues. Paxata's cluster and edit functionality uses NLP algorithms such as ngram, fingerprinting and metphone to automatically discover clusters of similar entity values and makes recommendations on combining them into a single golden record. In contrast to the classic rule-based approach, Paxata is continuously learning and evolving to automatically address known and unknown data quality issues. Enterprise-grade: Finally, as issues are uncovered and fixed, every change should be captured and tracked to address information security and governance requirements and give bankers the accurate and trustworthy information they need to confidently run the business. Paxata provides auditing, lineage, versioning, recording and reordering capabilities to track, undo and redo changes, operationalize and provide context on steps taken to prepare data. In conclusion, accurate and actionable information is paramount to success in banking and Paxata is the fastest path to turn raw data into reliable information. You can learn more about Paxata here. FREE TRIAL DataRobot Data Prep Interactively explore, combine, and shape diverse datasets into data ready for machine learning and AI applications Try now for free About the author Before we proceed, let's revisit the difference between those three terms. Data is facts. It is the raw, untouched data that is captured. Information is data and facts that have been somehow transformed, for example, via aggregation or categorization. Typically, information is what is visualized and included in reports. An insight is derived from that information, usually through analytics. Insights should consider the context of the problem/question at hand, and then draw conclusions, which will lead to decisions and actions. At each stage in this process, we must question both the validity and relevance of what is being shared. First, we need to critically appraise it to discern whether we can trust it to be a fact or insight. Then, we need to evaluate it to discern whether its relevance to the problem or question at hand. When we don't do this – and when we don't apply healthy skepticism and critical thinking at each stage in the process – we can end up making poor decisions.During the initial stage when you are looking at just raw data, there are situations in which that data is incorrect. For example, in 1492, Christopher Columbus sailed from Europe across the Atlantic Ocean to find an alternative route to Asia. But, Columbus relied on the erroneous calculations of several geographers from conflicting sources and eras to chart his route. In addition, Columbus did not convert Arabic miles used by one of the geographers to Roman miles, leading him to grossly underestimate the expanse separating the continents. This bad data led Columbus to land in the Americas and not Asia. Similarly, during the stage where you are organizing data into information, the information can also be incorrect. Perhaps the wrong transformation was used or the wrong categorization was applied. Potentially the information is accurate, but the definition of what the information is trying to show is misleading. For example, does everyone have the same interpretation of what the term "profit" means if the information is showing average monthly profit? Is it gross profit, net profit or perhaps something else? When the person sharing this inaccurate or misleading information does not know it is inaccurate or misleading, it is called misinformation. When the data and information have intentionally been shared knowing its inaccurate or misleading, it is called disinformation. Both misinformation and disinformation can lead someone to draw incorrect conclusions. If the data and information is validated as accurate and is relevant to the problem/question you are trying to address, you will then move on to the next stage and try to come up with insights. During this stage, it is equally important – and many times more challenging – to apply healthy skepticism before believing the insights to be true and acting on them. There are many reasons why insights may be either inaccurate or irrelevant, but here are six of the most common ones: Looking for trends where there may not be any Looking at correlations when there are not any Misunderstanding the results from an inferential statistic Incorrect mental models Looking at a symptom and not the root cause Tunnel vision/lack of innovationEveryone is susceptible to misinformation/disinformation and false insights, because we are all susceptible to cognitive self-bias, which causes us to think we have the right information and to think we have the right insights when, in fact, we do not. We all have emotions, and typically are dealing with information overload, combined with a lack of time and attention, leaving us unable to complete everything needed. These can lead us to avoid challenging our assumptions and mental models.Like how scientists use the scientific method to apply rigorous skepticism to their observations, it is important to apply the same rigorous skepticism by critically questioning and challenging all data, information and insights before treating them as truths. To help with this, below is a list of questions that you can ask related to each stage in the process to help ensure you are working with quality insights that you can act on. Want to know more about how to reach high-quality insights that drive impactful decisions and help your organization achieve success? Check out our Data Literacy Global Studies and Insights webpage and download a copy of the "Data Literacy: The Upskilling Evolution" report, which Qlik developed in partnership with The Future Labs.