Hopf, Konstantin; Joshi, Mayur; Shollo, Arisa; Stelmaszak, Marta

Data-Based Craft : How Data Scientists Craft Their Data, Models, and Products

Date of secondary publication: 14.01.2026

Accepted Manuscript (Postprint), Bookpart

Persistent identifier: urn:nbn:de:bvb:473-irb-112513x

Primary publication

Hopf, Konstantin; Joshi, Mayur; Shollo, Arisa; Stelmaszak, Marta (2025): Data-Based Craft : How Data Scientists Craft Their Data, Models, and Products, in: Vern L. Glaser, Christine Moser, Deborah A. Anderson, P. Devereaux Jennings (Ed.), Algorithmic Organizing, Leeds: Emerald Publishing, pp. 19–43, doi: 10.1108/s0733-558x20250000095002.

DATA-BASED CRAFT: HOW DATA SCIENTISTS CRAFT THEIR DATA, MODELS, AND PRODUCTS

Konstantin Hopf,[a] Mayur P. Joshi,[b1] Arisa Shollo,[c] and Marta Stelmaszak[d2]
[a]*Chair of Information Systems and Energy Efficient Systems, University of Bamberg*
[b]*Telfer School of Management, University of Ottawa*
[c]*Department of Digitalization, Copenhagen Business School*
[d]*Isenberg School of Management, University of Massachusetts Amherst*

This is the author accepted manuscript.
The final version of record is available via
https://doi.org/10.1108/S0733-558X20250000095002

ABSTRACT

In this study, we examine the work of data scientists, members of an emerging technical occupation, through the lens of craft. Drawing on 65 in-depth interviews with data scientists, we show that their work cannot be adequately explained by the human-machine configurations characterized in the existing literature on craft in technical occupations, which primarily focuses on crafting products using ready-to-use tools and ready-to-be-processed materials. Instead, we find that data scientists craft not only their products, but also their tools and materials, often in an iterative and non-linear fashion. This distinct approach entails a unique human-machine-data configuration that we refer to as data-based craft, which stems from the unique nature of digital data and learning algorithms that data scientists simultaneously craft and use. This study advances our understanding of craft in the digital age by highlighting the need to reconceptualize human-machine relationships in data-intensive occupations.

Keywords

Data scientists, craft, data-intensive occupations, changing nature of work

[1] Corresponding author; mayur.joshi@telfer.uottawa.ca
[2] All authors contributed equally and are listed alphabetically

INTRODUCTION

With the surge in technological advancements in the postindustrial economy, technical occupations are attaining a prominent place in the division of labor (Adler, 2007; Barley, 1996) as new occupations tend to coalesce around emerging technologies (Bechky, 2020). Scholars agree that the members of various technical occupations inherently engage in some form of craft, which can be explained by how work is configured in human-machine relationships (Kroezen et al., 2021; Sennett, 2009). Human-machine configurations reflect some form of division of labor in which humans leverage machines (or tools) to process materials to craft products. Accordingly, various technical occupations can be conceived as distinct types of craft—e.g., structural engineering as traditional craft, the work of computer technicians as industrialized craft, and the work of lab technicians as technical craft—as per the typology of craft proposed by Kroezen et al. (2021) based on their comprehensive review of the literature.

Although past research has largely focused on products as the key objects of craft and treated machines as ready-to-use and materials as ready-to-be-processed (Bailey et al., 2010; Barley, 1996; Kroezen & Heugens, 2019; Orr, 1996), we encountered a new technical occupation of data science—one of the most hyped technical occupations of recent times (Davenport & Patil 2012, 2022)—whose members not only craft the products but also their materials as well as tools. In particular, we found that data scientists not only craft their products, but also spend considerable time crafting their tools[3] (i.e., algorithms and models) and materials (i.e., data). Because they are not merely users of machines and materials, their work cannot be explained by the human-machine relationships theorized in the craft literature.

To explain this empirical surprise, we turn to the emerging literature on data science,

[3] We use "machine" to abstractly refer to technology and "tool" to refer to more concrete elements of craft.

data-intensive technologies, and machine learning (ML) algorithms in the information systems and computer science disciplines (Cao, 2019; Donoho, 2017; Joshi et al., 2021; Shollo et al., 2022; Vaast & Pinsonneault, 2021). We highlight two features of data science work that make it a distinct technical occupation. First, data scientists engage with a unique kind of materials—digital data—that differ substantially from the materials used in other technical occupations. Digital data are malleable, constantly changing, and lack the physical stability of materials used in other forms of craft (Aaltonen et al., 2021; Jones, 2019; Kallinikos, 1995; Kallinikos et al., 2013). Second, data scientists' tools—e.g., ML algorithms—also are unique. These tools (i.e., algorithms) need to be extensively trained with materials (i.e., data) (e.g., Shollo et al., 2022; van den Broek et al., 2021), before they can be employed to craft products, and they are never fully finished (Waardenburg & Huysman, 2022). The unique nature of the materials and tools used in data science and their mutual dependence orchestrated by data scientists warrants a thorough examination of their everyday work practices to understand how they craft their products in contemporary organizations.

We examine these dynamics in a qualitative study based on 65 in-depth interviews with data scientists. We find that data scientists not only craft under-defined and incomplete data science products for business customers (e.g., dynamic prediction dashboards or reports), but also actively craft tools by developing algorithmic models, and materials by generating and shaping data. We identify several practices associated with crafting materials (i.e., generating materials for a specific purpose, making materials processable, and assigning meaning to materials), crafting tools (i.e., exploring effective tools, tuning tools, and testing tools), and crafting products (i.e., envisioning, pitching, and cultivating products).

Based on these findings, we develop a model of *data-based craft* that explains the work of data scientists and can be extended to professionals in other data-intensive occupations, such as economists, engineers, scientists, and knowledge workers (Dougherty &

Dunne, 2012). Data-based work entails crafting all elements (materials, tools, and products), and does not follow a linear progression; rather, it requires multiple iterations of crafting each element in no fixed order. Importantly, data are not the only materials; the tools and products can themselves become materials for subsequent projects. Data-based craft is underpinned by the malleability of data, the autonomy of tools, and the inherent incompleteness of crafted products, shifting the focus of data-based craftworkers from creating final products to oscillating between crafting each of the elements.

CRAFT IN THE WORK OF TECHNICAL OCCUPATIONS

Technical occupations have been surging since industrialization, as increased work complexity and specialization necessitate technical assistance (Adler, 2007). A considerable stream of literature conceives the work of technical occupations (among others) as craft (Sennett, 2009; also see Kroezen et al., 2021). We use the occupations of structural engineering, computer technicians, and lab technicians as illustrative examples to characterize craft work in technical occupations as summarized in Table 1 (columns A to C).

-----Insert Table 1 about here-----

In technical occupations, humans play a vital role in making the work crafty. For instance, in the craft of software engineering, software engineers are a key distinguishing factor that makes it a work of craft (when compared to the work of producing industrially manufactured goods). The emphasis is on “a humanistic approach to work that prioritizes human engagement over machine control” (Kroezen et al., 2021, p. 503). Human involvement makes crafted products unique because “in contrast to automated processes of production where there is a high level of certainty and standardization about the outcome, the objects of craftwork cannot be accurately predetermined” (Bell et al., 2018, p. 1). Humans make the work crafty by processing raw materials while interacting with machines (or tools) to create products. There are three noteworthy features of craft: The products that humans create, the materials they use, and the machines they work with.

First, the key focus of literature is on crafting the products. The products are crafted by combining various materials and using tools in novel and meticulous ways (Becker, 1978; Kroezen et al., 2021; Rowley, 1997). In other words, the expertise of technical occupations lies in being able to process materials using available tools to create products that are valuable and meaningful to clients. These products tend to be representations (in technicians' work) or creations (in structural engineering) (Bailey et al., 2010; Bailey & Barley, 2011).

Second, the machines that humans use in technical occupations range from simple artifacts (e.g., pencil) to complex tools (e.g., specialized drawing software). Irrespective of their complexity, the tools are usually static and ready-to-use for a specific functional purpose, and their functioning is deterministic (Barley, 1996; Kroezen & Heugens, 2019; Orr, 1996). For instance, Barley (1996) described how medical technicians interact with machines to craft their products (i.e., representations such as assays or descriptions of samples). In other words, most technical occupations depict some augmentation of human and machine skills in making products, accepting machines as part of their craft, without needing to modify them extensively.

Third, the materials used to craft products tend to be either physical entities or their representations, which follow the rules of natural sciences and are bound by their physical attributes. For instance, blood tissue (material for medical technicians) has identifiable properties based on knowledge from the medical, chemical, and biological sciences (Barley, 1996; Huising, 2014; Nelsen & Barley, 1997). Similarly, building materials in structural engineering have properties that follow the rules of physics and chemistry.

Although all technical occupations reflect the three features of craft explained above, each occupation is unique and reflects an idiosyncratic configuration of humans and machines (see Kroezen et al., 2021 for a review of various craft configurations).[4] For instance, the occupation of structural engineering (e.g., Bailey & Barley, 2011) represents a human-machine configuration akin to traditional craft, entailing substantive human engagement. Structural engineers traditionally rely on manual skills and relatively simple tools (paper, pencil) or rule-based software packages in creating engineering drawings. Another example is computer technicians, who work as brokers to protect clients from engaging with the technical world (e.g., Barley, 1996). Considerable work is done by machines, so technicians' engagement is limited to maintaining them. This configuration between humans and machines is akin to other industrialized craft occupations. Yet not all technicians are the same. Medical and lab technicians work with machines, acting as buffers to help medical professionals understand and engage with physical entities that are not accessible without machines. As such their work is deeply intertwined with machines akin to other craft occupations that Kroezen et al. (2021) refer to as "technical crafts."

In this study, we set out to examine craft in data science as a new technical occupation that coalesced around intelligent tools (Bailey & Barley, 2020; Barley & Beane, 2020; Bechky, 2020; Vaast & Pinsonneault, 2021). We found that the work of data scientists cannot be fully explained by any of the human-machine configurations that are reflected in other technical occupations previously examined in the literature (namely traditional craft as evident in structural engineering, industrialized craft as evident in the work of computer technicians, and technical craft as evident in the work of medical and lab technicians), as we illustrate in Table 1. In particular, we found that data scientists not only craft their products (like other technical occupations) but also craft their machines (i.e., algorithms and models)

[4] Other craft configurations, creative craft and pure craft, are equally important in the craft literature. We do not discuss these two configurations, as they are less relevant to the technical occupations that constitute the focus of this study. In addition, it is worth noting that technical occupations are not limited to the configuration of "technical craft" as demonstrated by Kroezen et al. (2021). Thus, we cover traditional as well as industrialized craft configurations.

as well as their raw materials (i.e., digital data). As a consequence, their work is not merely about the scale that hinges between humans and machines as evident in other craft configurations (see Kroezen et al. 2021), but more substantively about the ongoing and iterative process of crafting the material, the tools, and the products simultaneously. In order to make sense of this puzzling discovery, we reviewed the traditional as well as emerging literature on data science (Davenport, 2018; Dhar, 2013; Fayyad et al., 1996; Hopf et al., 2023; Joshi et al., 2021; Stelmaszak, 2022; Tukey, 1962; Vaast & Pinnensault, 2021) and interpreted our findings in the light of insights from this literature. We uncovered two features of data science that, we believe, contribute to making the work of data scientists unique, which cannot be explained by the human-machine configurations previously theorized in the literature on craft.

First, data scientists engage with materials that are very different from those used in other technical occupations. For most other types of craft studied this far, materials tend to be physical entities or at most their representations. In contrast, digital data are the materials used in data science. Such data are far from being “raw” (i.e., pre-existing in organizational reality) and “ready to use” as inputs to generate insights (Gitelman, 2013). Data are malleable, constantly changing, and lack the physical stability of materials used in other forms of craft (Aaltonen et al., 2021; Jones, 2019; Kallinikos, 1995; Kallinikos et al., 2013). For instance, data collected from traces of social media interactions or engagement with websites are representations (as is the case with traditional materials) as well as constructions of social behaviors (as opposed to physical entities). Such data lend themselves to reconfiguration, editing, and reprogramming (Kallinikos et al., 2013). Moreover, their value and usefulness reside in their semantics—that is, the meaning they carry with respect to the real-world entities they aim to capture (Alaimo & Kallinikos, 2021).

Second, data scientists interact with unique types of machines (e.g., learning algorithms, data pipelines) which need to be calibrated, trained, and tuned to perform well, yet often non-deterministically, on data which are constantly changing. In other types of technical craft (e.g., software engineering) tools may be chosen, adapted, or even fashioned from scratch (e.g., by writing code to match a pre-defined specification), but they promise stable, repeatable functioning leading to predictable outcomes. In contrast, data science involves prototyping or experimenting with different statistical approaches as the desired product emerges from the programming task (Kerry & Myers, 2017). Unlike other forms of craft, data science products are not always developed based on fixed and identified functional needs, and are constantly updated and changed as they are used, blurring the distinction between crafting tools and using products (Waardenburg & Huysman, 2022).

In summary, even though data science is a technical occupation, it has some characteristics that are not usually visible in other technical occupations (as highlighted in column D of Table 1). The differences lie in the inherent nature of the materials and tools used by data scientists, as well as the products they craft. Due to the idiosyncrasies of data science, extant literature on craft in technical occupations offers limited guidance on the question: *How do data scientists work with machines to craft their products?* Given the prevalence of data science in contemporary organizations, this question becomes increasingly important. In answering this question, we advance a model of data-based craft that represents a new configuration of data scientists, machines, and data that makes it possible for data scientists to create the desired products.

METHODS

Our qualitative study uncovers craft practices in data science work, a quintessential example of data-intensive technical occupations. We interviewed 65 data scientists in 25 globally distributed organizations ranging from high-tech to traditional firms in long-standing industries. All authors were involved in data collection and analysis. We independently found strong evidence of craft, for example, in data scientists' skills (e.g., skill mastery, abstract expertise) and attitudes (e.g., dedication, exploration), both of which are typical of craft (Kroezen et al., 2021). Following this revelatory discovery of craft in the unexpected setting of data science, we began to reanalyze our data in light of the conceptualization of craft in organization and management scholarship, as we describe in more detail below.

Data Collection

Adopting an inductive approach to data collection (Miles & Huberman, 1994), each member of the author team conducted interviews with at least 10 informants affiliated with data science teams. Around half of the interviews were conducted with members of data science teams in four organizations (one in India, one in the U.S., and two in Central Europe). Other informants were recruited in the spirit of maximum variation sampling (Patton, 2002). The goal of this sampling approach was to capture different perceptions and narratives about common issues within larger teams and varied individual perspectives from many firms. Informants had various levels of education (bachelor's: 9; master's: 26; PhD: 18; did not disclose: 12) in diverse fields, including statistics, mathematics, computer science, and data science. They reported work experience in data science between 1 and over 10 years (5.29 on average), and 10 were team leaders. See Table 2 for more details about the informants.

----Insert Table 2 about here-----

Interviews followed a semi-structured guide that was flexibly adjusted over time based on informants' accounts (Gioia et al., 2013). The interviews focused on what, how (Barley, 1996), and why (Eyal, 2013) data scientists do what they do. We grounded the interviews in informants' own experiences, asking about concrete data science projects, both past and present. Using project-specific examples, participants shared narratives about the practices and processes followed while generating insights from data and sharing insights with business clients in their attempt to generate value and facilitate organizational decision-making. This helped us keep the interviews rooted in actual events and settings, yielding insights about informants' perspectives (Gioia et al., 2013; Patton, 2002). All the interviews were recorded (with explicit consent), lasted 55.2 minutes, on average, and were transcribed verbatim in their original language (English: 38; German: 17; English combined with Hindi: 10). This yielded 59.25 hours of recordings and 1,051 pages of transcribed text.

Data Analysis

We engaged in iterative data analysis and coding, drawing on grounded theoretic methods (Glaser & Strauss, 1967). Our work consisted of three rounds of analysis.[5] In the first round, we assigned codes based on informants' own words, thereby giving "extraordinary voice" to them as "knowledgeable agents" (Gioia et al., 2013, p. 26). Each member of the author team conducted open coding (Glaser, 1978) focusing on how data scientists do their work. This resulted in codes such as "experimenting with data," "exploring datasets," "configuring the algorithms," and "revising models accordingly in an iterative fashion." A surprising finding emerged from this initial analysis: data science was far from a mechanistic and standardized approach of employing abstract expertise. Rather, we found that it entailed active involvement of data scientists in generating insights from data. At this point, we turned to the literature on craft, especially the conceptualization of craft as an approach to work, to help us make sense of our initial findings.

In the second round of analysis, we adopted the craft lens, seeking to identify specific practices associated with enactment of craft. As a sensitizing instrument for describing craft practices, we used the characteristics of craft outlined by Kroezen et al. (2021) and the descriptions of materials, tools, and products of craft identified from our literature review. At this point, we re-read and re-coded the interviews through the lens of craft. We generated

[5] The data were transcribed before analysis, which also included translation of non-English interviews into English.

craft related first-order codes (Wiesche et al., 2017) by labeling the data scientists' practices. Comparing and contrasting the labels surfaced core categories that we identified as *sub-practices*, such as "storytelling," "identifying the right audience," and "reshaping data." We then consolidated these core categories into higher order themes, the *craft practices* of data science work, such as "envisioning products," "pitching products," and "cultivating products," among others.

Then, we classified craft practices into broad *craft activities* involved in data science. Although we identified that data scientists craft their products much like workers in other technical occupations in the extant literature on craft, we uncovered that they also dedicate significant efforts to crafting their materials and tools. Thus, craft activities in relation to all three elements of craft (materials, tools, and products) emerged as a transparently observable, yet unexpected finding.

In the third round, we reread the data multiple times as we analyzed and synthesized the findings. We also compared our emergent findings with the literature on craft in technical occupations (e.g., Barley, 1996). This comparative analysis helped us to identify the specific characteristics of data science work that differ from other technical occupations, enabling us to isolate specific reasons why data scientists craft their materials and tools, in addition to their products. Subsequently, we identified overarching themes associated with the activities and craft practices of data scientists. These cross-cutting themes captured the iterative nature of these activities and practices and enabled us to theorize and conceptualize a model of data-based craft.

FINDINGS

Our findings show that data scientists craft all elements of their work, including their materials (i.e., data), tools (e.g., models, algorithms, program code), and products (e.g., reports, insights, dashboards). Table 3 offers a summary of the three craft activities data scientists perform in their work, the specific craft practices associated with each activity, as well as how and why these craft practices take place. Below, we explain each activity—*what* they craft—and the underlying practices—*how* they do it.[6] Although we present these activities and practices sequentially, in reality they are highly iterative and intertwined.

-----Insert Table 3 about here-----

Crafting Materials: Data Science Inputs

Digital data constitute the materials data scientists use as inputs in other activities. Digital data, both structured and unstructured, often rapidly circulate in large quantities within and between various computers, servers, and data warehouses in organizations. Data scientists work with a variety of data, such as customers' financial transactions, and purchase and social media behaviors, employees' email exchanges, and athletes' performance scores, among many others. Data, the materials used in data science work, typically do not exist in ready-to-use forms, as they are malleable, changing, and frequently not pre-existing, as we explore below. We find that data scientists craft materials before using them to craft products through three practices: *generating idiosyncratic materials*, *refining materials*, and *assigning meaning to materials*.

Generating Idiosyncratic Materials

In many cases, when data scientists receive requests to solve business problems, business clients neither provide nor identify datasets that data scientists can use. Data scientists are responsible for identifying relevant sources of data but need to reach out to other stakeholders (I17). They are frequently faced with situations where the required data do not exist, making it necessary to "find the gaps when there are no data and finding solutions

[6] In presenting our findings, we include select quotes that vividly illustrate the activities and practices of data scientists. Due to space constraints, the full coding table is included in a separate supplementary document and is available from authors upon request.

to get more information on topics where there are gaps" (I52).[7] In other cases, relevant materials exist in organizational information systems as outputs of previous data science projects. Because most data were crafted for different purposes, they may not be aligned with the needs of the specific project at hand. Below, we explain how data scientists generate materials by identifying, creating, and remolding data.

When specific datasets have not been identified for the tasks at hand, data scientists seek to identify data sources, both internal and external to the organization. A data scientist working for publishers explained:

> Where do we get the data from? That is a big issue. Partly, we crawl social networks, buy media control data, tapping Amazon, the New York Times and Spiegel's bestseller list. Again, bringing the whole thing to a homogeneous level and making it uniform [is a huge effort]. (I18)

This often involves drawing from extensive experience and mastery, deploying analytical thinking, talking to employees from other parts of the organization, and combining variables and data structures in complex ways.

When the required data do not exist, data scientists create data from existing digital objects circulating within organizations, e.g., emails, transactions, or communication platforms. They do so by imagining what types of data will help them solve the problem at hand: "you have to do a lot of creative thinking about the data that we want to create, to insert into our model" (I51). This does not mean that they conjure up data, but that they carefully identify obtainable data or sometimes construct simulation models for aspects that cannot be measured objectively.

In some cases, data scientists remold data science outputs from other projects to avoid "reinventing the wheel" (I18). Data scientists use their expertise in remolding those materials so they can be used for a given project. Beyond changing the format or selecting the right

[7] To protect their identities, all informants are assigned unique reference numbers starting with the letter I.

variables, this involves imbuing data with the appropriate semantic value for a given project, which may differ from previous attributions.

Refining Materials

Sometimes relevant materials may already exist in databases, but data scientists still need to refine them to make them processable: “if we have data issues … we need those to be cleaned up” (I61). Often, data issues arise because they are collected from numerous different systems and databases. In addition, the need to make data processable is constant, as new data are continuously being generated by organizational information systems in line with changing organizational realities, so data scientists create “data pipelines” (I18). As such, changes are reflected in the data, they are constantly subject to “drift” (I12)—that is, changes in data collection, generation, and storage processes that cause the data to become less relevant and outdated, necessitating constant refinement.

Data scientists engage in the initial pre-processing of data, also called data cleaning, which may include transforming data points into appropriate formats, removing outliers, and imputing missing values. One of the key activities is “feature engineering” (I36, I51): deciding which data are relevant, and how to combine or transform them into features that are meaningful for solving a given problem. While the initial pre-processing of data may initially look like a set of straightforward technical operations, it often requires data scientists to draw on expertise mastery of skills, but also a dose of creativity to turn data into materials that can be usefully and productively used in modeling.

Refining materials involves constantly processing incoming data, which is usually one of the most time-consuming tasks that data scientists report. A data scientist who works for an analytics vendor complained about huge differences between data exports for the same project in “multiple iterations. Like we do this for one year, and we do it once again, and then users think you just run it again. That didn’t happen. The company gave us different data [which needed to be made processable]" (I08). Changes to the underlying processes that result in data drift call for repeated crafting of the materials to ensure that data do not lose relevance and become obsolete for the project at hand. This requires data scientists to be carefully attuned to the materials so they can sense and recognize the smallest of changes, which can lead to amplified consequences in later modeling stages.

Assigning Meaning to Materials

A lot of work data scientists do to generate new materials and make them processable is based on the meanings they assign to them. The materials they work with are semantic in nature, as they are imbued with specific meanings. Hence, data only become materials when data scientists assign them meaning. Data scientists see this practice as foundational to understanding the data and what they represent. Furthermore, meanings need to be assigned to data because contemporary organizations are constantly evolving and changing entities whose primary function is not to generate data. Rather, data are byproducts: "This is an organism, and it's something that is moving, and it produces breadcrumbs of data. The data did not produce the organization; the organization produced the data" (I48). These byproducts need to be assigned a purpose before they can be used.

Data scientists often assign meaning to materials by explicating the semantics of data, drawing on their expertise. A data scientist working for multiple energy retailers explained:

> That's also part of the exploratory thing, [to clarify] a lot of questions about the data, like: What is this? Why is it like this? What does this mean? These are not even data dictionary problems—more like an understanding of how the client's internal processes work. (I08)

This emphasizes that embodied expertise can be obtained through dedication to constant development and learning.

Data scientists also assign meaning to data by working collaboratively. One informant described this as a "tandem model in which the department works together with me, having the computer science/mathematics/statistics lens" (I03). This practice allows data scientists to access often implicit knowledge, a type of “tribal knowledge” (I52) that is best sourced and decoded in collaboration with others.

To attune themselves to the various meanings of data in organizational contexts, data scientists frequently resort to getting their hands dirty. They observe not only internal organizational processes but also interactions with the customers: “Doing analytics while sitting in the corporate office will not make you a good data scientist. For that, you need to get your hands dirty by going out in the real world” (I29). By witnessing customer interactions first-hand, data scientists can assign meaning (e.g., customer behavior and motivation) to the materials they craft.

Crafting Tools: Data Science Models

Data scientists use tools, such as ML algorithms, to process materials and craft products. For example, they use classification algorithms that can analyze and model credit card transactions and identify which ones are likely to be fraudulent. The behavior of such tools, however, is often “unpredictable” or non-deterministic,[8] and tends to generate different outputs with each instantiation, even with only minor changes to input data or parameters. In addition, some of these tools (e.g., learning algorithms) can evolve because with every iteration in training, they can autonomously update their weights, which changes a model’s functioning and impacts the resulting predictions. Thus, data scientists need to carefully craft their tools rather than simply use them. Unlike in other forms of craft, data scientists must engage in substantial and continuous tool crafting because their outputs need to be calibrated and tuned, and they may consistently produce different or unexpected results until they are properly “trained,” and perhaps even afterwards. In fact, crafting tools is an important part of data scientists’ work. During the interviews, we identified three practices through which data

---

[8] ML algorithms are often described as “stochastic” (i.e., involving uncertainty because of statistical noise, random errors, incomplete sampling, and generalization) and make use of randomness or probabilistic decisions (Goodfellow et al., 2016).

scientists craft their tools: *exploring effective tools*, *tuning tools*, and *testing tools*.

*Exploring Effective Tool*s

Data scientists need to find tools that are suitable for the problem at hand. Increasingly, data scientists are specializing in the application of specific tools and losing familiarity with the breadth of available solutions. Moreover, the number of models and approaches is constantly growing as the field evolves. Thus, solving a business problem requires extensive and constant research. In addition, data science is an emerging field. Exploring effective tools helps data scientists discover and maintain the distinctiveness of their role and professional identity. This clear identity helps organizations recognize situations where "they really just need a data analyst [rather than a data scientist]" (I44).

To cope with the extreme complexity and variety of data science toolkits, data scientists often focus on tool mastery. They consciously specialize in the use of a specific tool and learn everything about it, knowing that other data scientists in the organization possess deep knowledge of other tools. More complex projects require several data scientists to collaborate:

> We have two people that are stronger than I in classic ML. And one of those is probably also pretty competent in some architecture, not necessarily in front-end, but in writing back-end code. And then finally, there is another person who is decent in modeling or kind of classic ML, [but] knows no front-end. (I54)

Unlike other forms of craft where a given craftworker is presumed to have the expertise to manage the entire process and deep familiarity with tools, data scientists typically need to collaborate with others and leverage different types of expertise.

Data scientists with specialized expertise cultivate relationships in multiple communities to keep up with constant developments in the field. They reach out to and maintain relationships with domain experts both inside and outside their organizations. Multiple data scientists described participating in regular local "meetups" (I02, I04) as an essential part of their job, as it enabled them to stay abreast of all developments in the field and remain familiar with the range of tools available. Unlike in more typical forms of craft, where independent craftworkers tend to build community with other independent craftworkers, data scientists are frequently embedded in (and accountable to) their organizations. By cultivating relationships in multiple communities, data scientists share novel ideas about the latest tools, their successes (and failures), and learn from those successes (and failures), all of which help them find the right tools.

Tuning Tools

Having researched and identified the right tools, data scientists obtain them either from open-source libraries (e.g., packages in Python) or through licensing arrangements with vendors. In either case, before these tools can be used to craft products, they need to be appropriately configured for a specific problem or project. For example, a data scientist tasked with solving a problem using a topic modeling approach (a text analytics method that identifies topics from a large corpus of documents) needs to tune the algorithm (e.g., Latent Dirichlet Allocation) by specifying various parameters such as the number of topics, the desired level of per-topic/per-word probabilities, and other hyperparameters. Tuning becomes even more complex when data scientists are working with a real-time stream of data that may be continuously changing (as discussed earlier), leading them to experiment with models and tuning approaches.

Experimenting with models involves changing parameters and variables to see which model yields the best results. The trial-and-error search for appropriate models and parameters requires a lot of time and resources from data scientists, but is seen as an essential part of their work: "Modeling is a creative process. We have to use our imagination to sharpen the model. When you start modeling, you start with 50–100 predictor variables, then you play around" (I23). Experimenting requires expertise, as the different approaches and steps in this practice are based on intuition and professional judgment. Experimentation at this stage is essential, as models can behave differently each time they are run. Although this autonomous learning makes ML algorithms (and others) so powerful, it also makes them challenging to build.

Data scientists may rely on solutions such as AutoML to automate the parameterization of tools. With help from such "tuners," data scientists can avoid repetitive and tedious experiments. However, our informants emphasized that the practice of tuning the tuners is essential:

> AutoML is totally awesome. But I do need to know what I feed to it. AutoML only does parameterizations. Great. I don't have to worry about clever setting of parameters anymore. But that's the second block, to get from 97.1 to 97.2 percent [performance]. But the really exciting question is, what do I put in AutoML? What has to come out? Where do I have to be careful? And what must I not have too much of and not too little of? (I09)

Data scientists remain responsible for setting the goal, specifying the parameters to be changed and value ranges, and closely overseeing these tools' attempts to improve models which may produce different outcomes every time, requiring them to re-tune the goals, parameters, and value ranges.

Testing Tools

Since data scientists craft their tools themselves, they need to be tested before they can be used to craft data science products. Testing tools is important, because data scientists need to validate whether they are doing what they were crafted to do—that is, whether they work appropriately on new real-world data they receive. In addition, ML methods seek to estimate causal relationships from data. That is, tools must be validated through real interventions to rule out spurious correlations. Thus, data scientists evaluate model performance and conduct real-world experiments.

Usually, only a portion of a dataset is used to train each model, enabling it to extract rules from the data. The remaining data are used to evaluate whether the model offers accurate predictions, as its performance cannot be known in advance. Performance on the test dataset can be assessed through quality metrics (e.g., accuracy, recall, precision) or the amount of time required for training or execution. This is often done using pre-packaged functions and libraries. Evaluating model performance may involve trade-offs, as more accurate models may be more time-consuming to run, or precision may require sacrificing recall. Drawing on their embodied expertise, data scientists intuit which performance metric is the most relevant for the problem, and select the best-performing model accordingly. Thus, a degree of individual craft goes into selecting performance metrics and engaging in model evaluation.

In-depth validation by conducting real-world experiments requires approval from organizational decision-makers. Such real-world experiments can be directly integrated into applications. A digital insurance company, for example, designed its "platform in such a way that it [enables data scientists to] run experiments—A/B tests, so to speak. [They can] take all possible variants of campaigns but also product features and select the statistically better option" (I19). This blurs the boundaries between crafting materials, tools, and products. A model that is validated by conducting real-world experiments becomes part of the product itself, and yet remains a tool in the hands of a data scientist who uses the results of real-world experiments to possibly generate new materials (data) to continuously improve the model and its outputs.

Crafting Products: Data Science Outputs

The core of data scientists' work entails crafting products—that is, model outputs and the insights derived from them. Data science products are designed for general use, taking the form of dynamic dashboards, slide decks, reports, and spreadsheets. Sometimes, these insights are packaged into intelligent systems that act based on model predictions and other business rules, or are embedded in business processes (e.g., robotic process automation). Data scientists build attractive, functional, and practical products through ongoing interactions with colleagues, business clients, data, as well as tools. Unlike in other forms of craft, these products are often changing, dynamic, and are never completely finished. Our findings show that data scientists craft products through the practices of *envisioning*, *pitching*, and *cultivating products*.

Envisioning Products

Most organizations do not clearly understand the exact nature of the problems they want data scientists to solve. Many business clients tend to think of data science as a kind of "magic" which can sense the issues automatically and solve them. Yet, there is a lot that data scientists need to do before they can solve a problem. One informant explained:

> [This is] a huge challenge for us, especially because a lot of them [businesspeople] can't visualize things or don't understand what's possible technically. We will often challenge them: Don't worry about what you think is possible. Ask us for what you think you need and will tell you if it's possible or not. (I47)

In other words, data scientists actively envision a product—i.e., create an image of the product and its functional purpose—before creating it. Envisioning provides the roadmap for crafting not only the final product, but also the associated materials and tools. This vision may dynamically change as materials and tools are crafted, unlike other forms of craft where pre-existing materials and tools are used to create the imagined product. At the same time, envisioning helps to create a collaborative atmosphere with business clients to reach a shared understanding of the product to be created. Envisioning unfolds as data scientists transform business problems into data-driven problems and explore data pertaining to specific business areas.

A common approach when data scientists envision a product involves transforming business problems into solvable data-driven problems. Often, "for analytics to come in, there needs to be a problem that business is facing" (I28). Business clients bring problems to data scientists, and sometimes data scientists ask business clients about their challenges. Through several interactions with business clients, data scientists draw on their knowledge of data, tools and a data-driven philosophy to iteratively narrow the scope of a project and envision a concrete product. In so doing, they transform business problems into solvable data-driven narratives and negotiate the amount of time and resources that customers will dedicate to projects.

Data scientists also envision products by experimenting with different datasets and identifying interesting patterns that could possibly be developed into solutions. Rather than starting with a business problem, their work often involves "exploring the unexplorable. Like a scout, right? And if we discover value, then we can essentially evaluate: Should we draw the attention of larger efforts? And that has been our vision" (I45). Data scientists thus explore possible business contributions as part of their identity based on their abstract expertise. This enables them to suggest potential solutions when problems arise. Apart from the satisfaction that data scientists derive from this practice, exploring new sources and types of data also benefits them when envisioning the product.

Pitching Products

Since data science products are inherently digital and intangible, they need to be presented in ways that enable business clients to recognize their value and use them in their decision-making. This nature of the end product makes a craft approach all the more important for data scientists:

> You can have the coolest AI thing in the entire world and go and sell it, [but] the audience does not want to hear how you pulled the algorithm together to develop it, and stuff like that. They want to know what outcome [business value] it will have and what it'll deliver [functional purpose]. (I48)

Given the nature of problem-solving processes in the organizations, and the digital and intangible nature of data science products, pitching involves storytelling, timing products to coincide with business priorities, and educating the audience.

Data scientists often pitch their products by engaging in storytelling. Highlighting the insights derived from products helps clients understand the value of model outputs. Data scientists consider the ability to develop compelling narratives as integral to their work. It is critical to make causal connections between their predictions and human behaviors (e.g., purchase decisions in case of marketing analytics) in a language that clients understand. By telling stories, data scientists emphasize the prescriptive nature of their predictions, enabling them to "earn a seat at the table" (I48) where decisions are made. This is unlike other types of craft, as data scientists must actively pitch their products for them to be actively used in organizations.

However, these pitching efforts may be in vain if data scientists do not time their products well to coincide with business priorities. Some data science products are not adopted because they do not match with current business priorities. To address this challenge, data scientists continue to work on exploratory projects, but do not pitch them until the time is right. A data scientist explained: "While there are some product portfolios which may not be a business priority at the moment, we try to develop exploratory projects for these products, and then sell them when they become hot" (I53).

Data scientists also support their pitching efforts by educating customers. Informants indicated that data science teams may train "citizen data scientists" (I13) in organizations to raise awareness about how data science can help them solve their problems. By educating clients, data scientists set expectations and open up the black box around advanced analytical tools: "It seems like AI is still a buzzword out there, so I don't really think they really understand what we do exactly" (I50).

Cultivating Products

Rather than concluding a journey, pitching often marks the start of a long-term endeavor for data scientists focused on keeping their products alive and preventing decay (e.g., due to data drift). A data scientist explained that "when we are satisfied with the result, we present it to the customer … we decide together on the cadence to refresh […] and we refresh it every month or quarter" (I62). Amid the uncertainties associated with data drift, changes in underlying processes and data, and external events (e.g., a pandemic), data scientists continuously cultivate their products to enable them to flourish.

A common approach used by data scientists to cultivate their products is competing against themselves, which frequently leads to the development of permanently incomplete artifacts. Our informants explained that they almost never stop working on a project. Artifacts sold as products are only temporary instantiations of their ongoing work. In other words, even after pitching or discarding a product, data scientists continue working on the model in the background. Similarly, for products that are successfully implemented, data scientists continuously strive to improve outputs. Competing with themselves helps them continually improve models and spot early signals of model decay.

Finally, data scientists refresh their products by feeding them new business data in the short term (e.g., monthly), retraining them in the medium term, and redesigning them to adapt to significant changes in business processes. For example, "sometimes the weights that the model learns are no longer up-to-date, we need to update the weights" (I62). Data scientists are aware of data and model drift; therefore, they build their products in ways that enable them to be refreshed. Data scientists never fully relinquish responsibility for their products to business clients; instead, they remain continuously involved in their cultivation.

DISCUSSION

In this study, we set out to investigate how data scientists work with machines to craft products. We found that data scientists craft not only products, but also the materials (e.g., digital data) and tools (e.g., algorithms and models) involved in their development. These three activities unfold in an intertwined and iterative manner (see Table 3). Based on these findings, we present a model of *data-based craft* (Figure 1) which provides a foundation for an emerging type of craft. Below, we explain the model of data-based craft, compare data-based craft with other types of craft, reinterpret the literature on data science from a craft perspective, and discuss boundary conditions and future research opportunities.

-----Insert Figure 1 about here-----

A Model of Data-based Craft

In data-based craft, occupational members adopt a *crafty approach to work crafting the products, but also the requisite materials and tools.* Materials (i.e., data) must be crafted because they typically are not in forms that can be directly used for a desired purpose. Tools (i.e., models and algorithms that are used to analyze data) also must be crafted to be applied to unique organizational settings, problems, and datasets. In addition, data science tools retain some autonomy, functioning as counterparts (Anthony et al., 2023) to data scientists, as their behavior might change with every iteration. This requires data scientists to be attuned and constantly attentive to how tools evolve, tweaking and adjusting them with every independent change in their composition. Finally, products need to be crafted, often in collaboration with internal or external clients. Importantly, data science products are never complete and are continuously updated as business processes, data, models, and client requirements change, often independently of each other.

Our proposed conceptual model shows that data-based craft *does not follow a linear progression* from generating materials, to developing tools, to creating products. Instead, data-based craft involves *oscillating* between all three practices—that is, crafting products may require data scientists to recraft materials (e.g., generate purposeful materials or assign meaning based on the newly-crafted requirements of the product). In turn, changes to materials may create a need to recraft tools. Data-based craft involves many iterations, not only within each practice, but also between practices. Further work is needed to investigate what triggers craftworkers in data-driven occupations to move back and forth between the different crafting activities, as well as to uncover whether different pathways in iterative work have implications for the products being crafted.

Finally, our model emphasizes that in data-based craft, *tools and products can become materials* in subsequent projects. For example, a model to detect attrition developed for one project may become the tool for a new project in which several models are brought together to create an ensemble predictive model. Likewise, data about employee emails that are the materials for one project may be delivered to an executive as a final product.

Differentiating Data-based Craft from Other Types of Craft

While we acknowledge several commonalities with other types of craft, data-based craft differs from them in a number of important ways. First and foremost, data-based craft is *defined by the nature of data*, materials which are malleable, constantly changing, and lack the stability of physical materials before they are appropriately crafted (Aaltonen et al., 2021; Kallinikos, 1995; Kallinikos et al., 2013). Our findings show that one of data scientists' main activities is crafting materials by generating, assigning meaning to, and refining data to make it usable for their purposes. In other types of craft, materials are assumed to be stable raw inputs (Bailey et al., 2010; Bailey & Barley, 2011), and thus require little work from craftworkers. In contexts involving some pre-processing of raw materials (Barley, 1996; Huising, 2014; Nelsen & Barley, 1997), modifications typically apply to their physical properties or representations. Data, however, are digital artifacts (Kallinikos et al., 2013) that have been imbued with technical, individual, and organizational meanings, but must be crafted to yield insights in specific contexts. This is very different from other forms of craft where the meaning and relevance of materials are tied to their physical characteristics, which are apparent to craftworkers (Bailey et al., 2010; Bailey & Barley, 2011). Crafting data as materials thus necessitates crafting associated tools and products. Data are always in flux (Aaltonen et al., 2021; Parmiggiani et al., 2022), impacting the stability of models and products based on them. If data are constantly changing, associated models need to be tuned and adapted, and products need to be cultivated accordingly.

Relatedly, data are semantic carriers—that is, they are valuable as materials because of what they represent (Aaltonen et al., 2021; Alaimo & Kallinikos, 2021). Data-based craft thus entails shaping the meaning of data at each stage of the process. Unlike in other forms of craft involving physical materials (Barley, 1996; Huising, 2014; Nelsen & Barley, 1997), data-based crafting plays out at the level of semiotics: Crafting is not about the data being processed, but the meaning derived from it. Unlike craft focused on mediated meaning, in data-based craft, the treatment of material aspects of data shapes their semantics, just as shaping their semantics influences their materiality. Our findings depict data science as an archetypical occupation in data-based craft, where the occupational label reflects the importance of the materials (data), unlike other technical occupations that tend to be labeled based on their products (e.g., software developers).

Second, data-based craft is *defined by the nature of algorithmic models as its tools*, which are complex, fragile, non-deterministic, and often opaque, with unclear boundaries between their creation and use (Waardenburg & Huysman, 2022). These tools tend to change depending on the materials (i.e., data) fed into them. Thus, when data pertain to subsequent product use, tool development never ends (Waardenburg & Huysman, 2022). These tools need to be trained (i.e., calibrated and tuned) to perform well, yet non-deterministically, on data that are constantly changing. Our findings highlight that due to the level of complexity and opacity of tools as well as their reliance on data for learning purposes, data scientists carefully craft the tools they use by exploring effective tools, tuning, and testing them. In other forms of craft, tools are usually static, ready-to-use for a specific functional purpose, and deterministic, even when they are complex (Barley, 1996; Kroezen & Heugens, 2019; Orr, 1996). Furthermore, the oscillation that our model depicts underlines the mutual dependence between data and models, in line with recent conceptualizations of actorhood (Gehman et al., 2024; Glaser et al., 2024) "as an assemblage of humans, algorithms, data, and artifacts" (Gehman et al., 2024, p. 2), and previous work highlighting the importance of data and algorithmic models in data science work (for a debate on the importance of data and algorithms, see Constantiou & Kallinikos, 2015; Markus, 2015). In sum, in data-based craft, tools are constantly in the making.

Third, craft traditionally has been conceived as an approach to work that is deployed by individuals (Rowley, 1997), with an emphasis on skill mastery, well-roundedness, and dedication (Kroezen et al., 2021). Typically, a craftworker is an individual who has mastered specific skills and handles all aspects of craftwork (materials, tools, products) from start to finish. Data-based craft, by contrast, is characterized by *distributed mastery*, wherein multiple specialized organizational members collectively contribute to craftwork as "a collaborative endeavor involving cross-disciplinary expertise" (Parmiggiani et al., 2022, p. 41). Because materials, tools, and products change and evolve at a fast pace, it is not possible for a single professional to be sufficiently masterful and well-rounded in all elements to effectively engage in craftwork from start to finish. In the case of data science, no single person can evenly and productively specialize in the work required to generate, process, and assign meaning to data while staying abreast of the newest developments in modeling and being able to envision, prototype, and sell products to customers. Thus, data-based craft relies on distributed mastery.

Finally, in data-based craft, *no product is ever finished*. Because of the characteristics of data and algorithmic models, the outputs of such work always require further improvements, maintenance, and refinement. Those who engage in this kind of work constantly revisit previous projects and products that have been handed over to clients. Due to their dynamic nature, crafted products are not concrete, finalized outputs like software applications (Martin, 2008; McBreen, 2001), but are in constant flux and can be reused, repackaged, and recrafted into something else. This changes the orientation of craft, as creating the product ceases to be the ultimate goal; instead, crafting at every stage becomes more salient. This shifts the focus of data-based craft from creating final products to productizing every element of their work.

Our model puts forward a novel type of craft that is configured in a tripartite relationship among humans, tools, and materials which manifests in the crafting of all three elements: materials, tools, and products (see Figure 1). Far from rejecting technology in pure craft (Beverland, 2005; Sikavica & Pozner, 2013; Weber et al., 2008) or assuming that mechanization can at best augment human characteristics (Piore & Sabel, 1984), our model of data-based craft signals that machines can be engines of craft that are essential for humans to engage in this type of work. In other words, both humans and machines can engage in craftwork. Thus, our conceptualization of data-based craft challenges “conventional oppositional dualisms of hand/machine making and digital technologies/craft” (Bell et al., 2021, p. 13).

Boundary Conditions and Future Directions

Our study has several boundary conditions that present opportunities for future research. First, our field-level study offers a good overview of the fundamental activities and practices of data science work, but does not offer context-specific explanations. Although we covered a considerable base by interviewing data scientists across geographies and industries, we had to rely on our limited interactions to construct their operational context. We did not observe their everyday work practices in their respective organizations. A natural extension of this work would be to examine how these activities and practices of data scientists’ craftwork change in idiosyncratic contexts and why.

Second, the type of data-based craft that we have proposed in this paper requires further examination. For example, investigating the presence and preconditions of the relationships between materials, tools, and products could be a focus for quantitative or configurational studies. Another area that merits more attention lies at the intersection of data-based craft and the development of learning algorithms. While this technology has the potential to automate many repetitive tasks and is already affecting data scientists' professional identity (Vaast & Pinsonneault, 2021), it remains unclear whether technological advancements will fundamentally alter the nature of data work, given that humans are the creators of artificial intelligence applications (Aleksander, 2017; Willcocks, 2020).

Finally, our interviews revealed that data scientists often start by crafting one element (e.g., products), and iteratively craft the other two (e.g., materials and tools). We sensed a potential pattern in terms of how they decide the order of these activities. Some informants mentioned that these choices depend on the organizational context. We have an idea of what these factors may look like, but since we did not collect systematic data to uncover this pattern, we cannot make claims about these dynamics. In future research, it may be fruitful to identify the unique configurations and sequences of crafting activities, the reasons for these sequences, and the consequences of following one pattern over another.

CONCLUSION

Many see digital automation as a threat to craft, the humanistic approach to work. Craftworkers may feel they "are becoming further divorced from the very thing that defines [them as] makers, crafters of things" (Langlands, 2019, p. 22). By empirically examining the work of data scientists and by laying the theoretical foundation for data-based craft, we have articulated a view in which "the machine is us, our processes, an aspect of our embodiment" (Haraway, 1990, p. 222). As work is increasingly becoming data-intensive, we have uncovered a new type of craft that allows data scientists to productively work with machines. Our observations and theorization deviate significantly from the narratives of automation, replacement, or even augmentation. We observed that a craft approach to data-based work (i.e., making materials, tools, and products) is driving technology implementations in organizations through new human-machine-material configurations. We discovered that data scientists craft not only products, but also associated materials and tools in an iterative manner. Our hope is that our empirical findings about data science work and our model of data-based craft set scholars on a new course focused on examining configurations of humans and machines as they collaboratively engage in craftwork to construct not only representations of organizational realities, but new realities of organizational life.

REFERENCES

Aaltonen, A., Alaimo, C. & Kallinikos, J. (2021). The making of data commodities: Data analytics as an embedded process. *Journal of Management Information Systems*, *38*(2), 401–429.

Adler, P. S. (2007). The future of critical management studies: A Paleo-Marxist critique of labour process theory. *Organization Studies*, *28*(9), 1313–1345.

Alaimo, C., & Kallinikos, J. (2021). Managing by data: Algorithmic categories and organizing. *Organization Studies*, *42*(9), 1385–1407.

Aleksander, I. (2017). Partners of humans: A realistic assessment of the role of robots in the foreseeable future. *Journal of Information Technology*, *32*(1), 1–9.

Bailey, D. E., & Barley, S. R. (2011). Teaching-learning ecologies: Mapping the environment to structure through action. *Organization Science*, *22*(1), 262–285.

Bailey, D. E., & Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, *30*(2), 100286.

Bailey, D. E., Leonardi, P. M., & Chong, J. (2010). Minding the gaps: Understanding technology interdependence and coordination in knowledge work. *Organization Science*, *21*(3), 713–730.

Barley, S. R. (1996). Technicians in the workplace: Ethnographic evidence for bringing work into organizational studies. *Administrative Science Quarterly*, *41*(3), 404–441.

Barley, S. R., & Beane, M. I. (2020). How should we study intelligent technologies' implications for work and employment? In S. R. Barley & M. I. Beane (Eds.), *Work and technological change* (pp. 69–115). Oxford University Press.

Bechky, B. A. (2020). Evaluative spillovers from technological change: The effects of "DNA envy" on occupational practices in forensic science. *Administrative Science Quarterly*, *65*(3), 606–643.

Becker, H. S. (1978). Arts and crafts. *American Journal of Sociology*, *83*(4), 862–889.

Bell, E., Dacin, M. T., & Toraldo, M. L. (2021). Craft imaginaries—past, present and future. *Organization Theory*, *2*(1), 2631787721991141.

Beverland, M. B. (2005). Crafting brand authenticity: The case of luxury wines. *Journal of Management Studies*, *42*(5), 1003–1029. h

Cao, L. (2019). Data science: Profession and education. *IEEE Intelligent Systems*, *34*(5), 35–44.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44–57.
Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766.
Dougherty, D., & Dunne, D. D. (2012). Digital science and knowledge boundaries in complex innovation. *Organization Science*, *23*(5), 1467–1484.
Eyal, G. (2013). For a sociology of expertise: The social origins of the autism epidemic. *American Journal of Sociology*, *118*(4), 863–907.
Gehman, J., Glaser, V. L., & Merritt, P. (2024). An assemblage perspective on hybrid agency: A commentary on Raisch and Fomina's "Combining Human and Artificial Intelligence". *Academy of Management Review*, amr.2024.0126.
Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31.
Gitelman, L. (2013). *'Raw data' is an oxymoron*. MIT Press.
Glaser, B. G. (1978). *Theoretical sensitivity*. Sociology Press.
Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Aldine Publishing.
Glaser, V. L., Sloan, J., & Gehman, J. (2024). Organizations as algorithms: A new metaphor for advancing management theory. *Journal of Management Studies*, *61*(6), 2748–2769.
Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
Haraway, D. (1990). A manifesto for cyborgs: Science, technology, and social feminism in the 1980's. In L. J. Nicholson (Ed.), *Feminism/postmodernism (thinking gender)* (pp. 190–233). Routledge.
Huising, R. (2014). The erosion of expert control through censure episodes. *Organization Science*, *25*(6), 1633–1661.
Jones, M. (2019). What we talk about when we talk about (big) data. *Journal of Strategic Information Systems*, *28*(1), 3–16.
Joshi, M. P., Su, N., Austin, R. D., & Sundaram, A. K. (2021, March 2). Why so many data science projects fail to deliver. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/why-so-many-data-science-projects-fail-to-deliver/
Kallinikos, J. (1995). The architecture of the invisible: Technology is representation. *Organization*, *2*(1), 117–140.
Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The ambivalent ontology of digital artifacts. *MIS Quarterly*, 357–370.
Kery, M. B., & Myers, B. A. (2017). Exploring exploratory programming. *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 25–29.
Kroezen, J., & Heugens, P. P. M. A. R. (2019). What is dead may never die: Institutional regeneration through logic reemergence in Dutch beer brewing. *Administrative Science Quarterly*, *64*(4), 976–1019.
Kroezen, J., Ravasi, D., Sasaki, I., Żebrowska, M., & Suddaby, R. (2021). Configurations of craft: Alternative models for organizing work. *Academy of Management Annals*, *15*(2), 502–536.
Langlands, A. (2019). *Cræft: An inquiry into the origins and true meaning of traditional crafts*. W. W. Norton & Co.
Markus, M. L. (2015). New games, new rules, new scoreboards: The potential consequences of big data. *Journal of Information Technology*, *30*(1), 58–59.
Martin, R. (2008). *Clean code: A handbook of Agile software craftsmanship* (1st ed.). Prentice Hall.

McBreen, P. (2001). *Software craftsmanship: The new imperative* (1st ed.). Addison-Wesley Professional.
Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.
Nelsen, B. J., & Barley, S. R. (1997). For love or money? Commodification and the construction of an occupational mandate. *Administrative Science Quarterly*, *42*(4), 619–653.
Orr, J. E. (1996). *Talking about machines: An ethnography of a modern job*. ILR Press/Cornell University Press.
Parmiggiani, E., Østerlie, T., & Almklov, P. G. (2022). In the backrooms of data science. *Journal of the Association for Information Systems*, *23*(1), 139–164.
Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Sage.
Piore, M. J., & Sabel, C. F. (1984). *The second industrial divide: Possibilities for prosperity*. Basic Books.
Rowley, S. (1997). Introduction. In Rowley, S. (Ed.), *Craft and contemporary theory*. Allen & Unwin.
Sennett, R. (2009). *The craftsman*. Yale University Press.
Shollo, A., Hopf, K., Thiess, T., & Müller, O. (2022). Shifting ML value creation mechanisms: A process model of ML value creation. *Journal of Strategic Information Systems*, *31*(3), 101734.
Sikavica, K., & Pozner, J.-E. (2013). Paradise sold: Resource partitioning and the organic movement in the US farming industry. *Organization Studies*, *34*(5–6), 623–651.
Vaast, E., & Pinsonneault, A. (2021). When digital technologies enable and threaten occupational identity: The delicate balancing act of data scientists. *MIS Quarterly*, *45*(3), 1087–1112.
van den Broek, E. Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly*, *45*(3), 1557–1580.
Waardenburg, L., & Huysman, M. (2022). From coexistence to co-creation: Blurring boundaries in the age of AI. *Information and Organization*, *32*(4), 100432.
Weber, K., Heinze, K. L., & DeSoucey, M. (2008). Forage for thought: Mobilizing codes in the movement for grass-fed meat and dairy products. *Administrative Science Quarterly*, *53*(3), 529–567.
Wiesche, M., Jurisch, M., Yetton, P., & Krcmar, H. (2017). Grounded theory methodology in information systems research. *MIS Quarterly*, *41*(3), 685–701.
Willcocks, L. (2020). Robo-apocalypse cancelled? Reframing the automation and future of work debate. *Journal of Information Technology*, *35*(4), 286–302.

Table 1. Key elements of craft across representative technical occupations

Aspect	A: Structural engineers	B: Computer technicians	C: Lab technicians	D: Data scientists
Representative studies	Bailey et al. (2010); Bailey & Barley (2011)	Orr (1996); Barley (1996); Kroezen & Heugens (2019)	Barley (1996), Nelsen & Barley (1997); Huising (2014)	This study
Product	Structural designs (models) of buildings	Diagnostic representations (e.g., interpretations of control panels on machines)	Diagnostic representations (e.g., assessments of patients' health conditions)	Dynamic dashboards, reports, web applications featuring the outputs of computational models
Functional purpose	Load bearing and safety	Seamless production of industrialized goods	Seamless delivery of services (e.g., patient conditions in healthcare)	Aid in decision-making, intelligent products or services
Materials	Representations of building materials in the physical realm with properties that follow the laws of natural sciences (e.g., iron, concrete)	N/A	Physical evidence (blood samples, heart rate, and other vital signs)	Malleable, reprogrammable data imbued with socially constructed meaning representing entities that exist in the physical realm
Tools	Paper, pencil, pocket calculators, and more recently, rule-based software packages	Diagnostic software packages	Tools for collecting and analyzing samples (stethoscopes, thermometers)	Non-deterministic, changing, and configurable algorithms and models
Human-machine configurations	Direct human engagement in making, traditionally relying on manual skills and relatively simple tools (paper, pencil), and recently, rule-based software packages, akin to "traditional craft"	Limited human engagement in making, as most work is done by machines; human role limited to maintaining machines akin to "industrialized craft"	Direct human engagement in making, deeply intertwined with machines akin to "technical craft"	This study

Table 2. Summary of interviews across organizations

Company (pseudonym)	Industry (category)	Company category	No. of interviews	Country (*=global)
C01	Professional services	Small	4	CH
C02	Manufacturing	Large	1	DE*
C03	Professional services	Small	1	DK
C04	Professional services	Medium	1	SG*
C05	Health	Large	1	DK*
C06	Finance/Insurance	Medium	2	DK
C07	Manufacturing	Small	1	DE
C08	Manufacturing	Small	1	DE
C09	Manufacturing	Medium	1	DK*
C10	Retail (incl. end-consumer)	Large	6	DK*
C11	Manufacturing	Medium	1	DE
C12	Retail (incl. end-consumer)	Medium	1	DE
C13	Retail (incl. end-consumer)	Medium	2	SE*
C14	Professional services	Small	1	DE
C15	Finance/Insurance	Medium	1	DE
C16	Retail (incl. end-consumer)	Large	2	US*
C17	Banking and financial services	Large	10	IN
C18	Professional services	Small	1	US
C19	Manufacturing	Large	2	DE*
C20	Manufacturing	Large	1	DE*
C21	Media	Medium	1	CH
C22	Retail (incl. end-consumer)	Large	1	DE
C23	Professional services	Large	20	US
C24	Professional services	Small	1	DE
C25	Retail (incl. end-consumer)	Medium	1	DE

Table 3. Summary of findings

Activities	Higher order themes: Practices in data science craft	Core categories: Sub-practices in data science craft	Reasons why data scientists need to engage in craft
Crafting materials: Data science inputs	Generating idiosyncratic materials based on embodied expertise and mastery of data science within the context	● Identifying data sources ● Creating data from other digital objects ● Remolding data from other data science products	● Absence of specific, indicated datasets to use ● Required data do not exist within the organization ● Underlying data are pre-crafted for different purposes
	Refining materials in response to constantly changing processes based on embodied expertise and mastery of data science	● Initial pre-processing of data ● Constant processing of data	● Dirty data (i.e., formatting and contents must be cleaned up) ● Data drift renders data less relevant and more obsolete
	Assigning meaning to materials based on engagement in the domain and collaboration with others	● Explicating the semantics of data ● Working in tandem to understand data ● Getting hands dirty to obtain embodied expertise	● Data as materials have semantic value that is not obvious from their characteristics ● The organization is a constantly evolving and changing entity; thus, meanings shift
Crafting tools: Data science models	Exploring effective tools requires skill mastery, communality and dedication to the role through continuous learning	● Leveraging distributed mastery ● Engaging in collaboration within and outside the organization	● Ever expanding number of potential models to be deployed forces specialization ● Working with specific tools as a matter of occupational identity
	Tuning tools involves constant experimentation and mastery of data science skills	● Experimenting with models ● Tuning the tuners	● Model parameters must be defined anew for every project ● Data format or structure may change ● Models may produce different results every time they are run
	Testing tools involves experimenting in real-world conditions and collaborating with other departments	● Evaluating performance ● Conducting real-world experiments	● Various performance metrics are available ● Evaluation on real-world data is needed ● Validating causality in detected correlations is required
Crafting products: Data science outputs	Envisioning products requires well-roundedness and collaboration with other organizational members	● Transforming business problems into data-driven solvable problems ● Exploring business area data	● Organizations struggle to exploit available data ● Data are ambiguous ● Business clients are not clear on what they want

Pitching products involves communicating the dedication of data scientists to their craft	● Storytelling ● Timing products ● Educating customers	● The digital and intangible nature of data science products makes them elusive to customers
Cultivating products requires dedication to ensure that products continue to perform well	● Competing against oneself ● Refreshing products	● Models decay over time ● Data drift renders data products less relevant and more obsolete

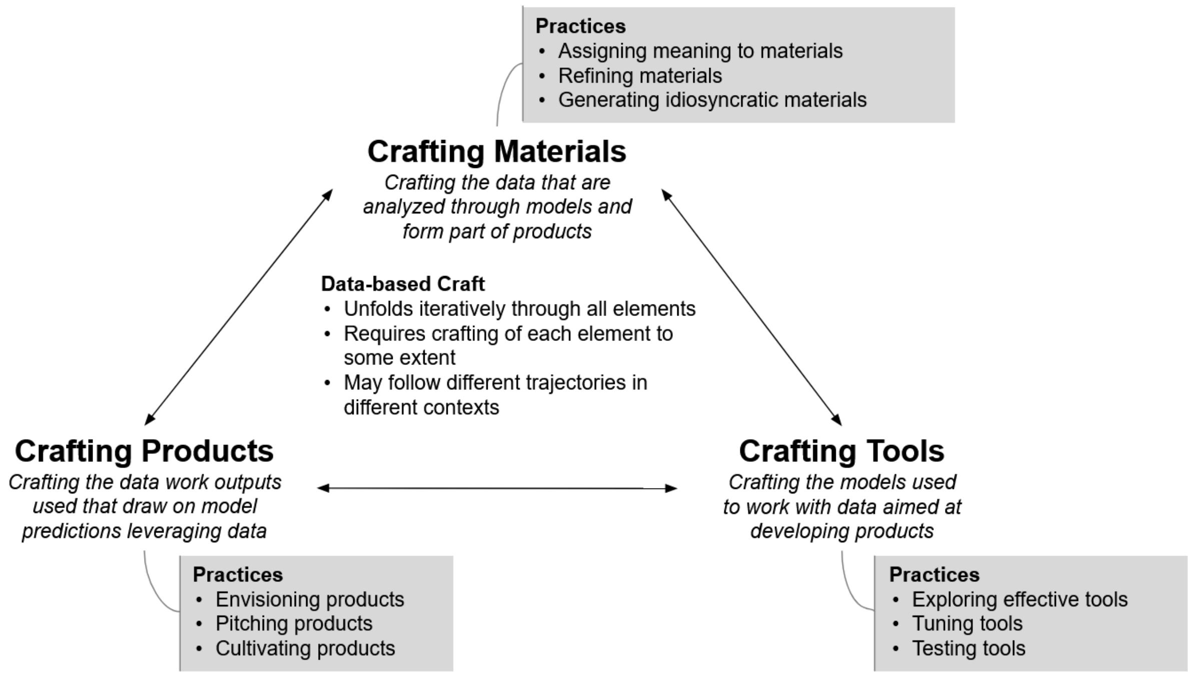

Figure 1. A model of data-based craft