# Understanding Human-Data Relationships: Data as Personhood

**Jed R. Brubaker**
Information Science
University of Colorado Boulder
jed.brubaker@colorado.edu

**Casey Fiesler**
Information Science
University of Colorado Boulder
casey.fiesler@colorado.edu

## Abstract

When does data belong *to* a person and when does data stand in *as* a person? The design and architecture of datasets are necessary selective and partial representations of the world, typically tailored to the computational needs of the systems from which they emerge. As a result, these systems often fail to reflect how end users think about online data. People's understandings of their relationships to the content they create and their digital traces—including their rights and the rights of others—has important implications for data science. In this half of our argument, we focus on the ways that personhood impacts how people orient to and relate to data. We assert that a human-centered approach to data science must consider and engage with the ways that "humanness" is operationalized within the data ecosystems that produce the datasets – large and small – that data science engages.

**Author Keywords**
Digital identity; personhood; Human-data relationships; User-generated content

**ACM Classification Keywords**
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

## Introduction

A central challenge in data science is understanding what the data being analyzed actually represents. With the explosion of social media, the ways data embody and represent people makes this challenge particularly pronounced. Social data does more than describe people. When people browse their friend's profiles, comment on photos, and interact online, they are interacting with digital doppelgangers designed as proxies that embody us, and in turn, stand in as us.

The ways that end-users experience this proxying is different than how it is experienced by data scientists. The design and architecture of datasets are necessary selective and partial representations of the world, typically tailored to the computational needs of the systems from which they emerge. As such, how data are structured and stored often does not map on to the ways that end-users understand their date. However, these data structures are central to how data scientists make sense of and analyze data, consequently shaping the types of analyses and claims that can emerge.

In the social media context, the "user account" is a dominating entity around which most user data – the online identity – is organized. The user account is used to represent a person to a computational system, however, in many cases it does not capture the ways people feel embodied online nor the relationships and rights a person may feel hey have to various data. As a result, we argue that understanding how people understand that their relationships to data – as contingent, contextual, and evolving – is important in a human-centered approach to data science.

## Background and Motivation

The way we interact with data, as well as what we capture in data, has changed dramatically over the last fifty years. However, the fundamental architectures developed for early operating systems still predominate the approaches taken to the design of user-data relationships. In contrast with legal approaches, technical systems have approached human-data relationships in terms of computer security, defining what users are associated with data as a way of operationalizing rights and permissions.

The concept of a user account was developed in the late 1950s as a time-sharing system [3], necessitating authentication mechanisms. The approach of linking users to content was developed for file systems in the early 1960s [14], and has subsequently been adopted by network systems, personal computers, and most online services. Typically, the user account (or "digital identity") is used to make claims about who it is, which are then translated into permissions to features or content by the security architecture. Beneath the surface, this user-data relationship is codified as content linked to a single user. This account is conventionally considered the "owner" of created or added content, meaning a granting of management rights. However, relationships around social media content tend to be more layered and complex than this might suggest. For example, while a Facebook user has some control over what *photos* they are tagged in, the architecture grants the user rights to the *tag* rather than photo itself. Rather than data architecture, Fiesler has found that people often understand their relationship to content in terms of social norms and ethical intuitions [6]. Likewise, people do not necessarily have good models about what rights a website has to their content [7].

There have been design efforts around the development of "user-centric" identity-management techniques. However, their aim has predominantly focused on projects such as as single-sign on that enables people to use a single authenticated identity across multiple systems [8,9,11]. While user-centric approaches consolidate the management of identities across various platforms into a system, there is increasing recognition that the design of these platforms shift the system administration burden to end-users if they want to control their online data [1], but often without the tools to do so. To this end, it becomes important that the complex data end-users are tasked with managing is structured and presented in ways that align with how people think about and understand their data.

Overtime, and particularly in the era of social media, the link between account and data has expanded beyond management to one of identification. Associations between users and data now indicate who the user *is* rather than only what the user can *do.* Brubaker's work on the management of post-mortem accounts and data. Survivors of the deceased have strongly diverging views on how best to manage post-mortem profiles [5]. The desire to modify post-mortem identities often conflicts with the desire to honor the choices the deceased had taken while alive and leave the profile untouched. The tension is compounded by the shared motivation to honor the memory of the deceased, as well as confusions around the legal rights survivors have [2,12,13]. Brubaker's proposed stewardship model is an alternative to inheritance for the management of post-mortem social media accounts that seeks to address this tension, but may not track well to existing legal frameworks [4] let alone the technical infrastructure discussed above.

Managing online data touches on both personhood (addressed here) and property (see Fiesler submission), but these two concepts become blurred by technical systems and data architecture: (1) Online data, especially social media, is representative of people, and on sites like Facebook they often become part of the overall online identity; and (2) User accounts allow people to take action within a system, but the "account" is itself an artifact. This conflation between personhood and property becomes especially pronounced when others may want to make use of these accounts or access the data associated with them, as in the case of death [10].

## Ongoing Work

Brubaker's prior work builds on the wealth of scholarship about impression management and self-presentation online to show how people's understandings of personhood can shift relationships, rights, and responsibilities to data. Distinction are blurred between data, data as a result of the person's actions, and the person themselves. However, as data increasingly moves beyond bounded profiles, or even bounded platforms, a human-centered account of personal data is vital to grounding data science efforts that seek to use these data to represent people and their lives.

In order to better understand the nuances of how personhood in data shifts across contexts, we are conducting a two-stage study on social media content. The first stage of our study involves scenario-based interviews designed to identify the practices, rights, and responsibilities that people believe they have or should have in relationship to digital data across social and technical contexts. An interview study that will also allow us to identify the attributes (of data, people, and relationships between them) that act as pivot points for people's beliefs and opinions. If our interview participant is tagged in a photo, we might ask about rights and practices related to this photo. We want to know both what she thinks the actual current state of rights is, as well as whether she thinks these should be different than her intuitions. Therefore, we might ask who can and should have the right to view, delete, modify the photo, etc. Through analysis, we will be able to identify salient attributes about data that influence how people approach the affordances they expect—for example, who created it, when it was created, how it was shared, or in what medium.

The second phase of our study involves a large-scale Facebook-based survey in which people are asked to articulate their relationship to specific pieces of social media data drawn from their own networks, the ways in which they feel represented by these data, and subsequently the rights the imagine they should have. The content presented to participants will be selected based on the various attributes identified during the interview portion of this study as a way of scaling qualitative findings to best understand in what data configurations personhood is most salient to end-users.

We have three goals motivating this work: First, to identify types of relationships between people and data that exist beyond the technical architecture and data structure made available to data scientists. Second, to identify when, why, and how personhood impacts how people relate to data, and subsequently when we need to think about data as a person rather than assets, records, or other traces associated with a person. Finally, this work sits in a broader analytical arc in which we ask how the operationalization of people as "users" constrains our analytical approaches to human subjectivity, and as a result, the kind of data analytics we perform.

## References

1. Gail-Joon Ahn, Mohamed Shehab, and Anna Squicciarini. 2011. Security and Privacy in Social Networks. *IEEE Internet Computing* 15, 3: 10–12.
2. Natalie M. Banta. 2014. Inherit the Cloud: The Role of Private Contracts in Distributing or Deleting Digital Assets at Death. *Fordham Law Review* 83.
3. RW Bemer. 1957. How to consider a computer. *Automatic Control*.
4. Jed R. Brubaker, Lynn S. Dombrowski, Anita M. Gilbert, Nafiri Kusumakaulika, and Gillian R. Hayes. 2014. Stewarding a legacy: responsibilities and relationships in the management of post-mortem data. *Proc. CHI '14.*
5. Jed R. Brubaker, Gillian R. Hayes, and Paul Dourish. 2013. Beyond the Grave: Facebook as a Site for the Expansion of Death and Mourning. *The Information Society* 29, 3: 152–163.
6. Casey Fiesler and Amy S. Bruckman. 2014. Remixers' understandings of fair use online. *Proc. CSCW 2014.*
7. Casey Fiesler, Cliff Lampe, and Amy S. Bruckman. 2016. Reality and Perception of Copyright Terms of Service for Online Content Creation. *Proc. CSCW 2016*.
8. Dick Hardt. 2006. OpenID Attribute Exchange 1.0 -- Draft 1.
9. Audun Jøsang and Simon Pope. 2005. User centric identity management. *AusCERT Asia Pacific Information Technology Security Conference*, June: 1–13.
10. Stephan Micklitz, Martin Ortlieb, and Jessica Staddon. 2013. "I hereby leave my email to...": Data Usage Control and the Digital Estate. *2013 IEEE Security and Privacy Workshops*, IEEE, 42–44.
11. D. Recordon and D. Reed. 2006. OpenID 2.0: a platform for user-centric identity management.
12. Suzanne B. Walsh. 2014. Coming Soon to a Legislature New You: Comprehensive State Law Governing Ficuciary Access to Digital Assets. *Charleston Law Review* 8: 429–449.
13. Ashley F. Watkins. 2014. Digtial Properties and Death: What Will Your Heirs Have Access to After

You Die? *Buffalo Law Review* 62: 194–235.

14. Martin H Weik. 1961. *A third survey of domestic electronic digital computing systems No. BRL-1115*. Aberdeen Proving Ground, MD.