

A framework for centering racial equity throughout the administrative data life cycle

Dr. Amy L. Hawn Nelson^{1*} and Sharon Zanti¹

Submission History

Submitted:	30/05/2020
Accepted:	12/08/2020
Published:	30/09/2020

¹University of Pennsylvania

Abstract

Introduction

Data integration by local and state governments is undertaken for the public good to support the interconnected needs of families and communities. Though data infrastructure is a powerful tool to support equity-oriented reforms, racial equity is rarely centered or prioritized as a core goal for data integration. This raises fundamental concerns, as integrated data increasingly provide the raw materials for evaluation, research, and risk modeling. Generally, institutions have not adequately examined and acknowledged structural bias in their history, or the ways in which data reflect systemic racial inequities in the development and administration of policies and programs. Meanwhile, civic data users and the public are rarely consulted in the development and use of data systems.

Objectives

This paper presents a framework and site-based examples of “Work in Action” that were collaboratively generated by a civic data stakeholder workgroup from across the U.S. in 2019–2020.

Methods

Purposive sampling was used to curate a diverse 15-person workgroup that used participatory action research and public deliberation to co-create a framework of best practices.

Results

This framework aims to support agencies seeking to acknowledge and compensate for the harms and bias baked into data and practice. It is organized across six stages of the administrative data life cycle—planning, data collection, data access, use of algorithms/statistical tools, analysis, and reporting and dissemination. For each stage, the framework includes positive and problematic practices for centering racial equity, with site-based examples of “Work in Action” from across the U.S. Using this framework, the workgroup then developed a Toolkit for Centering Racial Equity Throughout Data Integration, a resource that has been broadly disseminated across the U.S.

Conclusions

Findings indicate that centering racial equity within data integration efforts is not a binary outcome, but rather a series of small steps towards more equitable practice. There are countless ways to center racial equity across the data life cycle, and this framework provides concrete strategies for organizations to begin to grow that work in practice.

Keywords

racial equity; public deliberation; data integration; participatory action research

Introduction

Data infrastructure + racial equity?

With trust in government and research at historic lows [1], efforts that rely exclusively on these institutions to “use data

to solve social problems” are unlikely to succeed. Too often government organizations and their research partners fail to identify and address issues of bias in data. Further, even if such issues are addressed, agencies are often ill equipped to repair trust and work towards justice in partnership with communities that have experienced harm.

*Corresponding Author:

Email Address: ahnelson@upenn.edu (Dr. Amy L. Hawn Nelson)

Cross-sector data sharing and integration enables the transformation of individual-level information into actionable intelligence that can be used to understand urgent and long-term community needs; improve services, systems, and practices; develop innovative policies and interventions; and, ultimately, build stronger communities [2]. Yet, the way cross-sector data are used can also reinforce legacies of racist policies and produce inequitable resource allocation, access, and outcomes [3, 4].

People living in poverty are often over-represented within government agency data systems, and disparate representation can cause disparate impact [5]. Furthermore, in the U.S., poverty is disproportionately experienced by Black, Indigenous, and people of color (BIPoC). Laws, policies, business rules, and narratives are affected by structural racism, which is the root cause of the racial disparities evident in system outcomes [4]. Such disparities demonstrate the consequences of structural racism: that, as a group, BIPoC in the United States have worse outcomes in many human service system measures regardless of socioeconomic status [6]. And yet, many agency solutions and data initiatives are largely disconnected from this root cause, and the “hunt for more data is [often] a barrier for acting on what we already know [7].”

Finding best practices

Since 2008, Actionable Intelligence for Social Policy (AISP) has coordinated a national network of local, state, and university partners (referred to as sites) that operate integrated data systems (IDS) to learn about service utilization patterns, understand risk and protective factors, track long-term outcomes, evaluate programs, inform policy, and drive practice improvements. Today, AISP’s network of established and developing IDS sites comprises over 50% of the U.S. population [8].

AISP has developed extensive knowledge and best practices for shared data infrastructure to support reuse of administrative data for research, evaluation, and policy analysis. However, when IDS sites or other data integration efforts—whether called data collaboratives, data trusts, data hubs, or some other term—sought guidance in using IDS for equity, justice, and centering community voice, there were few resources to offer or exemplars to replicate. This research seeks to fill that lacuna.

This paper describes the results of a collaborative workgroup process that formally began in January 2019, with the overarching goal of better understanding and documenting best practices for administrative data reuse that centers racial equity. The workgroup engaged in participatory action research and public deliberation to co-create a framework of best practices.

Purposive sampling was used to curate a 15-person workgroup, representing diverse perspectives across race, gender, geography, and profession. A range of participants—including community organizers who have successfully challenged data integration efforts and longtime government administrators at the state and local level—were intentionally brought together in person and virtually over 18 months. The workgroup developed research questions and frameworks, collected data, and, ultimately, co-created and disseminated a Toolkit that provides guidance for centering racial equity throughout human service data use and integration [9].

The collaboratively generated Toolkit is organized with guidance and examples across the data life cycle. Each stage includes examples of positive and problematic practices as well as examples of “Work in Action” from sites across the U.S. that are currently working on centering racial equity within their agency’s data use [9]. This process is ongoing as the Toolkit is an evolving compendium of applications, examples, and lessons.

Methods

Workgroup recruitment

The primary goal during the recruitment process was to create a collaborative workgroup of civic data stakeholders and users, including institutional members, community representatives, and bridges between community and institutions (e.g., service providers, evaluators). Initial participant invitation/outreach was determined by utilizing published reports and national contacts from civic data networks.

The workgroup was carefully curated to ensure diversity of lived experience, demographics, training, focus of work, and geography. Invitations were issued based upon 1) experience with administrative data reuse and interest in equity, 2) interest in supporting best practice development, and 3) ability to commit to long-term participation. Each participant was offered an honorarium, and all travel expenses were covered.

Notably, workgroup members also brought diverse perspectives to lived experiences of administrative data reuse. For instance, several members had spent their careers in government using administrative data to support service provision and policy development, while others identified themselves as community organizers actively engaged in preventing unethical reuse of administrative data. Structural diversity was essential to the workgroup process as it exposed workgroup members to new sources of knowledge and increased the overall value of knowledge shared [10]. Over the course of 14 months, workgroup members engaged in participatory action research and public deliberation.

Participatory action research

Participatory action research (PAR) is a form of cooperative inquiry that honors the expertise of those historically without power (e.g. research subjects) and with power (e.g. academic researchers). PAR is a process of empowerment and self-actualization of expertise of those involved in the praxis, rather than the theory [11]. This approach was important because this inquiry’s purpose was to generate best practices, not “best theories.” Thus, workgroup members engaged with questions collaboratively, introspectively evaluated their own practices, and critically analyzed power within relationships. They engaged in a reflective process “directly linked to action, influenced by understanding of history, culture, and local context, and embedded in social relationships [12].”

Various data collection and documentation methods were used throughout the workgroup process, including detailed note-taking during two in-person meetings, regular email engagement, ongoing “check in” surveys, virtual conferencing,

and extensive collaborative document development. Each Toolkit section was collaboratively generated and approved by all workgroup members, with critical elements decided during two public deliberation events.

Public deliberation

Public deliberation methodology requires that a group of participants—in this case, the workgroup—consider diverse perspectives over a series of discussions in order to come to a reasoned decision [13]. The purpose is for participants to reach agreement on collective statements and recommendations that accommodate multiple perspectives [14].

The first public deliberation event was held with workgroup members in July 2019 and led by a workgroup member with extensive facilitation experience. The structure of this six-hour meeting was collaboratively generated in advance. Multiple workgroup members presented content to support framing of relevant topics, and throughout the process members reflected on their thoughts, beliefs, and actions based upon this framing. Example questions included:

1. Where are you situated within the field? What guides your thinking about where you place yourself?
2. What have been your most influential experiences and resources in shaping your thinking about data access and use?
3. What key concepts and terms need to be identified and defined when thinking about data sharing, integration, and equity?
4. At the end of this workgroup, what needs to exist that doesn't exist right now? And the corollary, what do we want to avoid making/doing?

Throughout the session the group developed shared agreements, language, and processes to guide collaboration. The main emphasis was on refining core questions guiding the Toolkit development work. While these were not agreed upon by the end of this initial session, the workgroup came to agreement in virtual follow up discussions. These core questions were:

1. How has your lead agency/collaborative acknowledged the importance of a racial equity lens and demonstrated a commitment to engage in data integration efforts that are legal, ethical, and center equity?
2. In your site context, how will the community and government learn, work, and be mutually accountable for using integrated data to inform, evaluate, and co-create structures, policies, practices, and narratives for equity?
3. What approaches will be most effective for integrated data infrastructure development and data use?

The second public deliberation, held in October 2019, included the same workgroup members, and sought to crystallize framing and determine Toolkit content. Reaching agreement on the operationalization of terms was a primary task.

An iterative process began with 100+ keywords, ultimately reduced to 15, including: administrative data reuse, bias, community, community engagement, consent, data, data ethics, data governance, data privacy, equity, racial equity, racism (individual, institutional, and structural), and social justice.

The workgroup also agreed to use an existing framework from the Government Alliance for Race Equity (GARE)—normalize, operationalize, organize—to structure findings and encourage action [15]. Workgroup members categorized necessary concepts to consider and processes to engage in order to effectively center racial equity. Examples of positive and problematic practices throughout the administrative data life cycle were generated from members' experience and/or observation of civic data use. Examples were then discussed, critiqued, deliberated, and agreed upon prior to inclusion in the findings.

Notably, all public deliberation outcomes were captured in detailed meeting notes and included in shared documents. All findings went through extensive checking by all workgroup members and additional external reviewers. The final Toolkit was also extensively reviewed and critiqued, by reviewers (who received an honorarium), both internal and external to the deliberative process. External reviewers were suggested and agreed upon by members of the workgroup based on content expertise and experience in this intersectional space of agency data access and use and racial justice. The final version of the Toolkit was reviewed and "approved" by all workgroup members prior to release.

Results

The PAR and public deliberation processes resulted in the development of a Toolkit for community-based and government agencies regularly reusing administrative data. The Toolkit includes guidance on encouraged and discouraged data access and use.

Broadly, the workgroup strongly encourages:

- Inclusive participatory governance around data access and use
- Social license for data access and use
- A developmental approach to data sharing and integration

And broadly discourages:

- Access to individual-level linked data
- Data use for enforcement or investigation actions against residents
- Use of predictive algorithms without determining responsibility, explainability, accuracy, auditability, and fairness [16]
- Use of linked data across institutions that have patterns of institutional racism, specifically, law enforcement, which may have committed significant racialized harm without sufficient safeguards in place

In addition, the Toolkit describes positive and problematic practices for centering racial equity across the six stages of the

Table 1: Examples of positive and problematic practices, planning

Positive Practices	Problematic Practices
Including diverse perspectives (e.g., community members with lived experiences and agency staff who understand the data) on planning committees	Using only token “representation” in agenda-setting, question creation, governance, or IRB review
Building capacity for researchers, administrators, and community participants to work together on agenda-setting	Using deadlines or grant deliverables as an excuse to rush or avoid authentic community engagement
Researching, understanding, and disseminating history of local policies, systems, and structures involved, including past harms and future opportunities	Using only historical administrative data to describe the problem, without a clear understanding of harmful policies and a co-created plan of action to improve outcomes
Lifting up research needs of the community to funders; helping shape funding strategy <i>with</i> funders to support community-driven research	Accepting grant/philanthropic funding for a project that is not a community priority or need

Table 2: Examples of positive and problematic practices, data collection

Positive Practices	Problematic Practices
Adhering to data management best practices to secure data as they are collected—specifically, with carefully considered, role-based access	Assuming that programmatic staff collecting data have training in data management and data security.
Including agency staff and community stakeholders in defining which data should be collected or reused	Inviting only researchers to identify data needs
Collecting only what is necessary to your context	Failing to consider which data carry an elevated risk of causing harm if redisclosed when determining which data to collect in your context (e.g., a housing program that collects resident HIV status)
Strong efforts to support metadata documentation, including key dimensions of metadata such as: <ul style="list-style-type: none"> description provenance technical specifications rights preservation citation 	Failure to clearly identify, explain, and document data integrity issues, including data that are: <ul style="list-style-type: none"> inaccurate undocumented unavailable incomplete inconsistent
Including qualitative stories to contextualize quantitative data	Allowing quantitative data to “speak for itself” without context or discussion

data life cycle and highlights current site-based work taking place at each stage, as outlined below:

Racial equity in planning

Planning includes all work to prepare for future stages, such as identifying stakeholders, convening work groups, articulating a mission or purpose for data integration, developing understanding of the local and historical context, creating ethical guidelines, and developing a project plan.

Broward County, Florida demonstrates how using PAR in planning can infuse racial equity throughout the data life cycle. The county’s data collaborative intentionally involves system participants in governance, research, evaluation, and solution

creation to address racial, economic, and social/spatial gaps between predominantly White researchers governing IDS and those using public services. In planning, Broward County is creating an IDS that supports sharing strengths-based stories about their community while using data to co-create system and policy improvements [9].

Racial equity in data collection

Administrative data are typically collected for operational purposes; therefore, potential risks for reuse appear in research, evaluation, and policy analysis. These data are vulnerable to biases, inaccuracies, and incomplete or missing data, and often include communities that are over-surveilled by governments.

An equity lens considers these inherent vulnerabilities in data collection and how they can be reduced or contextualized appropriately.

The Allegheny County, Pennsylvania Department of Human Service’s initiative to collect sexual orientation, gender identity, and gender expression (SOGIE) data for youth in the child welfare system validates an intersectional approach to centering equity in data collection. This added data dimension allows for disaggregated analyses by SOGIE and race, which promotes visibility of these intersectional identities and supports equitable service provision and resource allocation. For this effort, the Department had to address privacy and

data security concerns surrounding youth SOGIE data, the implications of sharing these data with external stakeholders, and the complexities and costs of updating IT systems for secure data collection. Additionally, the Department engaged with IT staff to ensure they knew the importance of these changes to mitigate harm during the design process [9].

Racial equity in data access

Using a multi-tiered approach to data access allows for clear delineation of both the practical and legal availability of administrative data. This includes an emphasis on privacy

Table 3: Examples of positive and problematic practices, data access

Positive Practices	Problematic Practices
Open Data	
Open data that have been identified as valuable through engagement with individuals represented within the data	Ongoing open data that is based upon problematic indexes or algorithms, with a history of discriminatory impact on communities (e.g., release of “teacher effectiveness scores” and “school report cards”)
Clear data release schedules and information on where to go and how to access data once they are released	Releasing data that can be re-identified (e.g., data released by small geographies may be identifiable by local residents)
Restricted Data	
Adhering to data management best practices for data access, including clear data destruction parameters (if applicable) following use Utmost care given to de-identification and anonymization of data prior to release	Assuming that data management best practices are being followed without explicit protocols and oversight in place Releasing data that can be re-identified (e.g., data that have not been properly anonymized or include aggregate or subgroup data without suppressing small cell sizes)
Accessible data request process with clear policies and procedures for submitting a request and how requests are evaluated	Unwillingness to release data, or limiting access to researchers or individuals with existing relationships
Unavailable Data	
Clear documentation of why data are unavailable (e.g., specific statute, legislation, data quality explanation, data are not digitized, undue burden in data preparation)	Refusal to release data when release is permissible and would not pose an undue burden

Table 4: Examples of positive and problematic practices, use of algorithms and statistical tools

Positive Practices	Problematic Practices
Involving diverse stakeholders in early conversations about the purpose of an algorithm prior to development and implementation Clearly identifying and communicating potential benefits and risks to stakeholders Human-led algorithm use (i.e., human(s) can override an algorithm at any point in the process)	Developing and implementing algorithms for human services without stakeholder involvement or alignment across multiple agencies Implementing an algorithm with no clear benefit to individuals included in the data Elevating algorithmic decision making over judgment of seasoned practitioners; no human involvement
Using “early warning” indicators to provide meaningful services and supports to clients	Using “early warning” indicators for increased surveillance, punitive action, monitoring, or “threat” amplification via a risk score

Table 5: Examples of positive and problematic practices, data analysis

Positive Practices	Problematic Practices
Using participatory research to bring multiple perspectives to the interpretation of the data	Describing outcomes without examining larger systems, policies, and social conditions that contribute to disparities in outcomes (e.g., poverty, housing segregation, access to education)
Engaging domain experts (e.g., agency staff, caseworkers) and methods experts (e.g., data scientists, statisticians) to ensure that the data model used is appropriate to examine the research questions in local context	Applying a “one size fits all” approach to analysis (i.e., what works in one place may not be appropriate elsewhere)
Correlating place to outcomes (e.g., overlaying redlining ¹ data to outcomes)	Leaving out the role of historical policies in the interpretation of findings
Using appropriate comparison groups to contextualize findings	Making default comparisons to White outcomes (e.g., assuming White outcomes are normative)
Disaggregating data and analyzing intersectional experiences (e.g., looking at race by gender)	Disregarding the individual or community context in the method of analysis and interpretation of results

preserving technical approaches and data access protections to prevent reidentification. Moreover, determining which data are open, restricted, or unavailable can have significant equity implications. For example, releasing data on school level scores that are merely a proxy for poverty may lead families to view a school as failing and opt out of local schools. This type of data release can disproportionately impact individuals and communities that do not have access to other schooling options. Alternatively, not releasing data necessary for addressing a community-based problem may also have disproportionate negative impact, such as not releasing use of force data from a local police department to better understand racial disparities.

Tulsa, Oklahoma’s Birth through Eight Strategy for Tulsa (BEST) data collaborative provides an example of balancing access to integrated data while protecting privacy and data security. The collaborative brought together data from 32 programs across local government, nonprofit, private sector, and philanthropic organizations in order to address race, equity, and service overlap challenges in the community. BEST piloted a platform utilizing privacy-preserving record linkage that supported data integration while keeping individual and organizational data private and secure. The platform’s use of cryptographic technology allows researchers to integrate data more quickly and at lower cost, while also enhancing privacy for individuals and organizations [9].

Racial equity in algorithms/statistical tools

Algorithms reflect the biases of those who create them and the data used in their processes; for this reason, algorithms cannot be race neutral. However, certain strategies can and should be used to ensure transparency, assess algorithmic bias, and determine potential positive and negative consequences of applying algorithms in practice. *Fairness, Accountability, and Transparency in Machine Learning* promotes five

guidelines—responsibility, explainability, accuracy, auditability, and fairness—that serve as a starting place to inform the development and use of algorithmic tools in ways that are accountable to the public [16]. It is recommended that those developing and using algorithms draft a *Social Impact Statement* that describes how these five principles (or others relevant to local context) will be operationalized in order to formalize a commitment to the public and to these principles.

In May 2018, New York City convened a task force to assess the use and proliferation of automated decision systems (ADS) across city services. Prior to the task force’s first public forum, four graduate students built a website, Automating NYC, designed to make ADS conversations more accessible to the community. They worked with city agencies to develop case studies across social services systems, adapted nontechnical activities to demonstrate algorithmic concepts, and incorporated individual stories to accompany technical explanations. The site also included practical action steps and allowed community members to ask informed questions about how ADS contribute to unjust systems, with the hope that future systems are built to benefit the community.

Racial equity in data analysis

A racial equity lens during data analysis incorporates individual, community, political, and historical contexts of race to inform analyses, conclusions, and recommendations. Solely relying on statistical outputs will not necessarily lead to insights without careful consideration of data quality, disaggregation, and statistical power. However, disaggregation is also a series of tradeoffs. Without disaggregating data by subgroup, analyses can unintentionally gloss over inequity and lead to invisible experiences. Alternatively, creating a subgroup may shift the focus to a population that is already over-surveilled. Given the complex series of decisions inherently involved in centering equity within analysis, iterative

¹Redlining refers to the systematic denial of services by private businesses as well as federal and local government agencies in the U.S. based upon race and ethnic categorizations. Most well known was the practice of denying mortgage loans for homeowners in “redlined” areas [17, 18].

Table 6: Examples of positive and problematic practices, reporting and dissemination

Positive Practices	Problematic Practices
Developing differentiated messaging for different audiences that considers the appropriate level of detail and technical jargon, language, length, format, etc.	Using intentionally dense language with low readability, especially for non-native language learners
Reporting results in an actionable form to improve the lives of those represented in the data (e.g., analyzing food purchase data to identify food deserts and guide development of grocery stores)	Reporting data that are not actionable or that are intended to be punitive (e.g., analyzing food purchase data to remove recipients from other public benefit programs)
Acknowledging structural racism or other harms to communities that are embedded in the data	Attempting to describe individual experiences with aggregate or “whole population” data without analyzing disparate impact based on race, gender, and other intersections of identity
Providing clear documentation of the data analysis process along with analytic files, so that others can reproduce the results	Obscuring the analytic approach in a way that limits reproducibility

work with strong participation from a variety of stakeholders is critical.

#ChangeFocusNYC, a PAR project born out of a partnership between New York City’s Administration of Children’s Services (ACS) and Department of Education, aims to understand experiences of NYC youth involved with multiple city agencies and recommend policies that could benefit them. Fifteen youth were chosen to partner with academics to design and implement #ChangeFocusNYC. Youth investigators participated in all research phases and were essential contributors during development of the analytic plan. Collaboratively generated answers to research questions are helping ACS work toward a system in which young people are continuously engaged in shaping the institutions that impact their lives [9].

Racial equity in reporting and dissemination

Reporting and dissemination involves strategic consideration of the audience and mode of dissemination that most effectively conveys the information. Reporting on data work can include written documents, infographics and data visualizations, website materials, press releases or news articles, and even speeches. Across these mediums, centering racial equity means paying attention to which data are highlighted and how they are framed, as well as the readability and accessibility of the communication method.

The City of Asheville, North Carolina created a story map, “Mapping Equity in Asheville,” which links racial demographics over time to location [19]. Major increases in population, tourism, and economic activity over the past decade have had unintended negative consequences for low-income residents and residents of color, leading to widespread gentrification and displacement. Publishing results of geospatial analyses online in a user-friendly format has allowed Asheville residents to better understand the connection between racialized policies and physical location, particularly in regard to redlining practices. The story map provides valuable information to

government and community members to inform policy, programming, and resource allocation.

Discussion

Overall, the results of this process underscore the notion that centering racial equity throughout data integration is not a single, discrete step, but rather an ongoing process that must be included at each stage of the data life cycle. Each stage presents new opportunities to apply a racial equity frame, as well as new challenges and considerations.

When we began planning this formal inquiry in 2018, we assumed we would find site-based exemplars of centering equity throughout data integration efforts internationally, and efforts focused specifically on racial equity within the U.S. Findings from our initial literature review and landscape scan were underwhelming. While the intersection of data use and racial equity has emerged into the mainstream of public discourse within the U.S. more recently, response from government agencies and data intermediaries has been minimal [20]. The intersection of racial equity and data infrastructure seems to be of interest in theory, but challenging to put into practice.

Our interviews and discussions with “leaders” in this space were interesting, in that the leading-edge sites, including Allegheny County, PA and Broward County, FL, were forthright in their stance that they did not view themselves as exemplars. Both sites were clear in their commitment to racial equity and the learning process, but still had room for growth in their practices. As described in the Toolkit, while there is strong practice across sites, the work is considered periphery, and not centered.

We anticipate several reasons for this, the most obvious being that awareness of and discussions related to race and equity continue to be challenging for most Americans, particularly White Americans, who make up the majority of civic data users (especially staff in the public administration and data science workforce) [21, 22]. Perhaps as BIPOC

become more represented in the civic data workforce, staff will become more comfortable navigating issues of race to improve practice. Relatedly, perhaps the growing documentation of racial disparities across all human service sector realms will demand that racial equity be centered across all activities, including administrative data reuse.

Another reason for the dearth of exemplars is resource-related. The majority of administrative data reuse occurs within or on behalf of government agencies, and scarcity of resources is commonplace. Funding data infrastructure is a perpetual challenge, and funding civic engagement, critical discourse, and related activities to support equitable data infrastructure is an additional, often deprioritized request. In many cases, the site-based examples of Work in Action were funded philanthropically, rather than through government allocations.

The bedrock of effective community involvement—trust—is established over years, and involves shifting policy, priorities, and practices based upon feedback that is bi-directional between agency and community. Yet, typical funding sources are dependent upon a clear scope of work that is product rather than process oriented. Products rarely lead to enduring staffing structures that support stability and trustworthiness. While the commitment to long-term community building is often present, the resources to support this work are not.

This lack of resources may reflect the difficulty of effectively communicating the challenges and opportunities of centering racial equity throughout the data life cycle. Executive leadership, policy makers, and researchers are inundated with competing priorities, and, while the workgroup findings indicate a clear ethical imperative for improved practice, an explanation compelling enough to lead to action remains elusive.

Perhaps the most promising finding that emerged during all stages of this project is that sites are eager to learn and improve their practices. Findings clearly indicate that there are actionable steps every site can take, right now, with whatever resources are available, to center racial equity. The Work in Action sites have generated more equitable power relations, innovative solutions, healing of some harms, and desire for more and deeper relationships across racialized hierarchies and segregated spaces.

Conclusion

We are at a pivotal moment, one in which the use of data is accelerating in both exciting and concerning ways. We have access to greater amounts of data than at any other point in our history, but practice lags behind, placing BIPoC, particularly those with intersecting marginalized identities, at the greatest risk of the “data-ification of injustice [7].”

Acknowledging history, harm, and the potentially negative implications of data integration for groups marginalized by inequitable systems is a key first step, but it is only a first step. To center racial equity, findings indicate that government agencies, researchers, and civic data users must center the voices, stories, expertise, and knowledge of communities in decision making. With inclusive participatory governance around data access and use, administrative data reuse can support collective action with shared power to improve

outcomes and harness data for social good. We must continue to build understanding and support for adopting positive practices by acknowledging the harm of current problematic practices throughout the data life cycle. To move these conversations forward and see positive equitable practices normed, resourced, and adopted, we must cultivate spaces where civic data users can come together and debate these nuanced topics in good faith to ensure ethical administrative data reuse.

Acknowledgments

We would like to acknowledge the funding support for this work from the Annie E. Casey Foundation, and the contributions of AISP, including Della Jenkins, Matthew Katz, Emily Berkowitz, TC Burnett, Dennis Culhane, and Kristen Smith.

We would also like to acknowledge the extensive contributions of the workgroup, including: Niiobli Armah, Bridget Blount, Angela Bluhm, Katy Collins, Sheila Dugan, Sue Gallagher, Laura Jones, Chris Kingsley, Ritika Sharma Kurup, Tamika Lewis, Rick Little, Tawana Petty, Raintry Salk, and Michelle Shevin.

We are also indebted to the contributions of sites featured as Work in Action towards centering racial equity, including: Allegheny County (PA), Department of Human Services, Office of Analytics, Technology, & Planning and Office of Equity & Inclusion; Automating.NYC; Birth through Eight Strategy for Tulsa (BEST); Children’s Services Council of Broward County (FL); City of Asheville (NC); City of Tacoma (WA); DataWorks NC; Kentucky Center for Statistics; Mecklenburg County (NC) Community Support Services; New York City Administration for Children’s Services & Youth Studies Programs at the CUNY School of Professional Studies; and Take Control Initiative (OK).

Statement on conflicts of interest

The authors declare they have no conflicts of interest.

Ethics statement

Collaboratively generated group norms were created and agreed to by all participants during the first deliberative session. All members of the workgroup and site-based contributors were volunteers and consented to participate in this collaborative process, and were financially compensated for their travel and received an honorarium for their participation where allowable.

Abbreviations and terms

Integrated Data System: IDS.

Black, Indigenous, People of Color: We intentionally use the acronym BIPoC as a term that seeks to recognize the unique experience of Black and Indigenous people within the United States. Use of this language supports pro-Blackness and Native visibility.

Open data: Data that can be shared openly, either at the aggregate or individual level, based on state and federal law. These data often exist in open data portals.

Restricted data: Data that can be shared, but only under specific circumstances with appropriate safeguards in place.

Structural racism: the normalization and legitimization of historical, cultural, institutional, and interpersonal dynamics that advantage Whites, “while producing cumulative and chronic adverse outcomes for People of Color.” Embedded within structural racism is institutional racism, the ways “policies and practices of organizations or parts of systems (schools, courts, transportation, etc.) create different outcomes for different racial groups.” From <https://www.racialequitytools.org/fundamentals/core-concepts/racism>

Unavailable data: Data that cannot or should not be shared, either because of state or federal law, lack of digital format, data quality, or other concerns.

References

- Rainie L, Keeter S, Perrin A. Trust and distrust in America [Internet]. Washington, DC : Pew Research Center; 2019. Available from: <https://www.people-press.org/2019/07/22/trust-and-distrust-in-america/>
- Fantuzzo J, Henderson C, Coe K, Culhane D. The integrated data system approach: A vehicle to more effective and efficient data-driven solutions in government [Internet]. Actionable Intelligence for Social Policy, University of Pennsylvania; 2017. Available from: https://1slo241vnt3j2dn45s1y90db-wpengine.netdna-ssl.com/wp-content/uploads/2017/09/The-IDS-Approach_Fantuzzo-et-al.-2017_Final.pdf
- Janssen M, Kuk G. The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*. 2016; 33(3): 371–377. <https://doi.org/10.1016/j.giq.2016.08.011>
- Gillborn D, Warmington P, Demack S. QuantCrit: education, policy, ‘Big Data’ and principles for a critical race theory of statistics. *Race Ethnicity and Education*. 2018;21(2):158–79. <https://doi.org/10.1080/13613324.2017.1377417>
- Barocas S, Selbst AD. Big data’s disparate impact. *Calif. L. Rev.* 2016;104:671. <http://doi.org/10.15779/Z38BG31>
- Hayes-Greene D, Love BP. The groundwater approach: Building a practical understanding of structural racism [Internet]. Greensboro, NC: The Racial Equity Institute; 2018. Available from: <https://static1.squarespace.com/static/578fa7e3d482e9af82f8f507/t/5c1b08a50ebbe8eec9f38d21/1545275564106/REI+Groundwater+Approach.pdf>
- Benjamin R. *Race after technology: Abolitionist tools for the new jim code*. Cambridge, MA: Polity Press; 2019.
- Actionable Intelligence for Social Policy. Network Sites [Internet]. University of Pennsylvania; 2020. Available from: <https://www.aisp.upenn.edu/network-sites-map/>
- Hawn Nelson A, Jenkins D, Zanti S, Katz M, Berkowitz E, Burnett TC, Culhane D, et al. A toolkit for centering racial equity throughout data integration. Actionable Intelligence for Social Policy, University of Pennsylvania; 2020. Available from: https://www.aisp.upenn.edu/aisp-Toolkit_5-27-20/
- Cummings JN. Work groups, structural diversity, and knowledge sharing in a global organization. *Management Science*. 2004;50(3):352–64. <https://doi.org/10.1287/mnsc.1030.0134>
- Gergen KJ. From mirroring to world-making: Research as future forming. *Journal for the Theory of Social Behaviour*. 2015;45(3):287–310. <https://doi.org/10.1111/jtsb.12075>
- Baum F, MacDougall C, Smith D. Participatory action research. *Journal of epidemiology and community health*. 2006;60(10):854. <https://doi.org/10.1136/jech.2004.028662>
- Abelson J. Using qualitative research methods to inform health policy: The case of public deliberation. *The SAGE handbook of qualitative methods in health research*. 2010:608–20. <https://doi.org/10.4135/9781446268247.n32>
- Teng J, Bentley C, Burgess MM, O’Doherty KC, McGrail KM. Sharing linked data sets for research: results from a deliberative public engagement event in British Columbia, Canada. *International Journal of Population Data Science*. 2019;4(1). <https://doi.org/10.23889/ijpds.v4i1.1103>
- Government Alliance on Race & Equity. GARE Communications Guide [Internet]. Local and Regional Government Alliance on Race & Equity; 2018. Available from: <https://www.racialequityalliance.org/wp-content/uploads/2018/05/1-052018-GARE-Comms-Guide-v1-1.pdf>
- Diakopoulos N, Friedler S, Arenas M, Barocas S, Hay M, Howe B, Jagadish HV, Unsworth K, Sahuguet A, Venkatasubramanian S, Wilson C. Principles for accountable algorithms and a social impact statement for algorithms [Internet]. FAT/ML; 2017. Available from: <https://www.fatml.org/resources/principles-for-accountable-algorithms>
- Rothstein R. *The color of law: A forgotten history of how our government segregated America*. New York, NY: Liveright Publishing; 2017.
- Nelson RK, Winling L, Marciano R, Connolly N, et al. Mapping inequality: Redlining in New Deal America [Internet]. University of Richmond; 2020. Available from: <https://dsl.richmond.edu/panorama/redlining/#loc=5/39.1/-94.58&text=intro>
- City of Asheville GIS. Mapping racial equity in Asheville, NC [Internet]. Asheville, NC; 2020. Available from: <https://avl.maps.arcgis.com/apps/Cascade/index.html?appid=10d222eb75854cba994b9a0083a40740/>

20. O'neil C. Weapons of math destruction: How big data increases inequality and threatens democracy. New York, NY: Broadway Books; 2016.
21. Campbell Jr G, Denes R, Morrison C, editors. Access denied: Race, ethnicity, and the scientific enterprise. New York, NY: Oxford University Press; 2000 May 4.
22. Eubanks V. Automating inequality: How high-tech tools profile, police, and punish the poor. New York, NY: St. Martin's Press; 2018 Jan 23.

