ETHICS Framework from Archives to Data

Version 1.0

Jamie Rogers - ORCID: 0000-0002-4849-114X
Rhia Rae - ORCID: 0000-0002-9154-8033
Sonia Santana Arroyo - ORCID: 0000-0001-7785-028X

Introduction

As cultural heritage organizations expand access to historical records and increasingly engage in the creation, transformation, and curation of data, the need for an ethics-driven community of practice has become more important. Projects involving archival materials now regularly include activities such as transcription, data curation, machine-readable collections, and computational analysis. These practices create new opportunities for discovery and research, but they also introduce new ethical responsibilities.

The ETHICS Framework was developed to support reflective and responsible decision-making throughout the lifecycle of archival data projects. The framework centers on principles of Engagement, Transparency, Harm Reduction, Integrity, Care, and Stewardship. It is designed as a practical tool to help librarians and archivists document decisions, identify risks, support accountability, and encourage ongoing reflection during the creation and use of data derived from archival materials.

This framework emerged through work involving archival data creation and curation, where ethical and interpretive complexities surfaced throughout the project lifecycle. Decisions about what to include, how to structure information, how to describe people and communities, how to manage uncertainty, and how to provide access all involved judgment and carried potential consequences.

We also found that ethical guidance existed across many interconnected fields (including ecology, sociology, anthropology, library science, and computer science), but these conversations were often distributed across different professional communities and publications. There was no single practical framework designed to help practitioners connect these ideas into an applied, iterative workflow for archival data projects. This framework was created to help bridge that gap.

The framework also recognizes that archival data work often takes place within environments shaped by institutional priorities, grant requirements, timelines, staffing limitations, technical infrastructure, assessment expectations, and systems of professional recognition. At the same time, projects may involve responsibilities to represent communities, cultural protocols, privacy concerns, or long-term stewardship obligations that do not always align neatly with institutional pressures or incentives. Rather than framing these realities as inherently oppositional, the framework encourages librarians and archivists to make these tensions visible, document tradeoffs transparently, and approach decision-making with care, accountability, and contextual awareness.

The ETHICS Framework is not intended to function as a compliance checklist or scoring mechanism. Instead, it is designed to support assessment, dialogue, documentation, and continuous improvement. Ethical considerations may shift over time as projects evolve, communities provide feedback, technologies change, or new risks emerge. The framework therefore treats ethics as an ongoing practice of reflection and stewardship rather than a fixed endpoint.

How to Use This Framework

The following guidelines and self-assessment framework are intended to support reflection, documentation, and continuous improvement. They are not a compliance checklist or an aggregate scoring tool. Each section represents a related set of ethical considerations and should be reviewed independently.

Not every consideration will apply to every project, collection, dataset, or data type. If a consideration does not apply, mark it as N/A and briefly note why. The goal is not to fully address every consideration, but to make decisions visible, intentional, and open to revision.

This framework can be used through multiple stages of a project:

- Before a project begins, use the framework to define purpose, scope, community considerations, potential risks, access needs, labor expectations, and documentation plans.
- During a project, return to the framework as materials are selected, transformed, described, reviewed, and shared. Use the notes column to document questions, decisions, uncertainties, and changes in approach.
- After a project is completed, use the framework to review what was done, identify unresolved concerns, document lessons learned, and plan for stewardship, revision, takedown, or future consultation.

Each parameter includes an ethical consideration and a space for notes. Dating notes can help track how thinking changes over the project lifecycle and can show the iterative nature of ethical decision-making.

Ethical considerations may also be interpreted through different ethical lenses. For example, the Markkula Center's ethical decision-making framework offers approaches such as rights, justice, common good, utilitarian, virtue, and care ethics.[1] Different lenses may lead project teams to notice different risks, responsibilities, or tradeoffs.

Evaluation Key

Fully Addresses Consideration
The consideration is meaningfully addressed. Document decisions, including ethical considerations and tradeoffs. Maintain ongoing review as part of long-term stewardship.

Largely Addresses Consideration
The consideration is mostly addressed. Note any areas of uncertainty or limitation. Revisit decisions as the project evolves and strengthen documentation where needed.

Partially Addresses Consideration
The consideration is moderately addressed. Reassess decisions in light of project goals, institutional values, and potential impacts on individuals and communities. Seek additional input, revise approaches, and address gaps before proceeding.

Does Not Address Consideration
The consideration is not meaningfully addressed or is absent from decision-making. Significant risks or ethical concerns may be present. Reevaluate approach and revisit underlying assumptions before continuing.

[1] University, Santa Clara. "A Framework for Ethical Decision Making." Accessed April 1, 2026. https://www.scu.edu/ethics/ethics-resources/a-framework-for-ethical-decision-making/.

Care for the People Doing the Work

Working with archival materials, especially those involving harm, marginalization, or sensitive histories, can be intellectually and emotionally demanding. Approach this work with care for yourself and others. Recognize limits, avoid overextension, and build in time for reflection, discussion, and, when needed, stepping away from the work.

At the same time, this work can create opportunities for more thoughtful, transparent, and accountable forms of data practice. Ethical data practices require not only responsible decisions about records and data, but also sustainable and humane conditions for the people doing the work. Practitioners are encouraged to reflect on where they were able to make meaningful changes, strengthen care and accountability within their projects, and support more responsible approaches to archival data creation and stewardship.

1 - Should We Do This Work?

1.1 - Purpose, Scope, and Alignment

This section focuses on clarifying the purpose of the project and the appropriateness of creating data from archival materials. It is intended to help you define why the work is being done, what will be included, and how decisions align with ethical responsibilities.

Before beginning data work, take time to:

- **Clarify the purpose of the project.**
  Describe what you are trying to achieve and how transforming materials into data supports your goals. Not all collections need to be datafied, and this step helps determine whether data creation adds meaningful value.
- **Assess organizational readiness and relationships.**
  Consider whether your organization has the knowledge, capacity, relationships, or community connections needed to responsibly carry out the work. Reflect on whether community partners, cultural experts, or collaborators should be involved to provide guidance, context, or accountability throughout the project.
- **Define the scope of the work.**
  Identify what materials will be included and what will be excluded. Consider whether the scope is appropriately limited to meet project goals without introducing unnecessary risk.
- **Evaluate the risk of extractive research[2] practices.**
  Consider whether the project treats records primarily as data resources without regard for the people and contexts they represent. Identify ways to prioritize care, accountability, and community benefit.

[2] Definition: A research paradigm rooted in imperial epistemology in which communities are positioned as objects of study rather than as contributors to knowledge. Cultural knowledge and lived experience are treated as raw materials to be collected, classified, and redistributed for the benefit of outside institutions, with little accountability to the communities studied. Contemporary forms persist when research locates the source of a problem within the community itself rather than in the broader structural and historical forces that shaped it.
Source: Smith, Linda Tuhiwai. Decolonizing Methodologies. 2nd ed. Bloomsbury Academic, 2012. https://www.bloomsbury.com/us/decolonizing-methodologies-9781786998125/. Chapters 1–4.

- **Assess alignment with institutional mission and public value.**
  Determine how the project supports institutional goals, scholarly inquiry, or community benefit. Document how this alignment informed decisions about selection and scope.
- **Review legal requirements as a baseline.**
  Identify relevant laws, regulations, and institutional policies (e.g., privacy, copyright, data protection, records restrictions) that apply to the materials. Record any permissions, restrictions, or risk assessments related to legal requirements. Legal permission does not necessarily mean ethical appropriateness. Where legal permissibility conflicts with ethical concerns, document those tensions and consider additional safeguards, restrictions, or alternative approaches.

How to Use This Section

- Use this section at the start of the project to determine whether and how to proceed.
- Revisit it if the scope changes, new materials are added, or new risks are identified.
- Use the Assessment Notes and Actionable Items column to document decisions, uncertainties, and changes over time.

Not all parameters will apply equally to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Appropriateness of selection	Clear rationale grounded in mission, ethics, and community benefit; harms carefully assessed.	Selection aligns with goals; basic harm assessment completed.	Rationale partially articulated; ethical impacts lightly considered.	Rationale unclear; selection risks harm; ethical issues unaddressed.	

Organizational readiness and community relationships	Organization demonstrates appropriate subject knowledge, community relationships, partnerships, or cultural expertise to responsibly support the work; collaborators or community partners involved where appropriate.	Organization demonstrates general awareness of community context and has some relevant relationships or expertise.	Limited community knowledge, expertise, or partnerships; responsibilities and relationship needs only partially considered.	No consideration of organizational readiness, community knowledge, or need for community relationships or collaboration.	
Avoidance of extractive research practices	Project explicitly evaluates risks of extractive data practices and demonstrates strategies that prioritize care, accountability, and community benefit.	Some consideration given to avoiding extractive approaches.	Limited awareness of extractive dynamics; mitigation strategies weak or unclear.	Project treats records primarily as data resources without consideration of extractive implications.	

Institutional mission alignment	Material selection clearly supports institutional mission, public service values, and scholarly or community benefit; alignment explicitly documented.	Selection broadly supports institutional priorities; alignment reasonably evident.	Alignment with institutional mission only loosely articulated.	No clear connection between selection and institutional mission or public benefit.	
Legal compliance	Applicable legal requirements are clearly identified, met across all project stages, and documented; the limits of legal compliance are recognized, and any tensions between legal permissibility and ethical considerations are addressed with appropriate safeguards.	Most relevant legal requirements are identified and generally met; legal considerations are documented at a general level, with some awareness of ethical limits and tensions.	Legal requirements are partially identified or inconsistently applied; documentation is limited, and ethical gaps beyond compliance are only minimally considered.	Legal requirements are not identified or addressed; compliance is unclear or absent, and ethical implications are not considered.	

2 - Who Is Involved and Affected?

2.1 - Communities and Stakeholders

This section focuses on identifying the people and communities connected to the materials and considering how they may be affected by the creation and use of data.

Archival records often reflect relationships, histories, and experiences of individuals and communities. In some cases, the data is not simply about a community but may be understood as connected to them in ways that involve shared history, identity, or responsibility.

As you work through this section, consider the following:

- **Identify communities and stakeholders.**
  Determine who is represented in the materials and who may be affected by the creation or use of the data. This may include individuals, descendant communities, organizations, or groups connected to the records.
- **Distinguish between record holders and affected communities.**
  The organization providing or holding the records may not be the same as the community represented in them. These groups may have different perspectives, interests, or concerns.
- **Consider opportunities for consultation and input.**
  Where appropriate, seek input on how materials are described, interpreted, and used. This may include terminology, narrative framing, or decisions about access.
- **Recognize different forms of engagement.**
  Engagement can range from consultation to collaboration. In some cases, communities may play a role in shaping decisions about description, access, or use of the data.
- **Consider governance and decision-making roles.**
  For some materials, particularly those connected to Indigenous communities, there may be established authorities (such as tribal governments or councils) whose input is important in guiding decisions about data use and access.
- **Support ongoing communication.**
  Engagement is not limited to a single point in time. Consider how feedback can be received and incorporated throughout the project lifecycle.

- **Respect preferences related to access and use.**
  Communities may have perspectives on how materials should be shared, described, or reused. Where possible, reflect these preferences in project decisions.
- **Plan for revision and response.**
  Consider how the project will respond if concerns are raised after release. This may include mechanisms for updating descriptions, adjusting access, or revisiting earlier decisions.

How to Use This Section

- Use this section early in the project to identify relevant communities and stakeholders.
- Revisit it when making decisions about description, access, and data use.
- Use the Assessment Notes and Actionable Items column to document who has been identified, what input has been received, and how it has informed decisions.

Not all projects will involve direct community engagement. If a parameter is not applicable, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Identification of affected communities and stakeholders	Communities represented in the materials are clearly identified; descendant, affected, and stakeholder groups are recognized and considered in	Relevant communities are generally identified; basic awareness of stakeholder relationships and implications is present.	Communities are only partially identified; stakeholder relationships and implications are unclear or underexplored.	Communities represented in the materials are not identified; no consideration of affected groups or stakeholders.	

	planning, representation, and engagement.				
Consultation and engagement practices	Communities meaningfully engaged through consultation, dialogue, or collaboration during project planning and implementation.	Consultation occurred at some stage of the project.	Limited or informal engagement with stakeholders.	No engagement with communities or stakeholders.	
Shared governance or decision-making authority	Communities share authority in decisions about description, access, terminology, or data use; governance structures clearly defined.	Community perspectives inform decisions but authority remains primarily institutional.	Consultation occurs but has limited influence on decision-making.	Community input not sought or incorporated into decisions.	

Respect for community preferences and restrictions	Community preferences regarding terminology, interpretation, access, and reuse are incorporated; mechanisms exist for restrictions, revision, or takedown when harm is identified.	Community preferences considered but only partially implemented.	Community concerns acknowledged but not systematically addressed.	Community preferences not incorporated into decision-making.	
Ongoing relationship and accountability	Engagement is iterative and sustained; mechanisms exist for feedback, correction, and future consultation.	Some ongoing communication with stakeholders.	Engagement limited to a single consultation or project phase.	Engagement is limited or primarily symbolic, with minimal influence on decision-making.	

3 - What Should Be Included?

3.1 - Data Selection and Exclusion

This section focuses on making intentional decisions about what information is included in the dataset and what is left out. Not all information in archival materials needs to be transformed into data. Careful selection helps reduce risk, avoid misinterpretation, and keep the dataset aligned with its purpose.

As you plan and carry out the work, consider the following:

- **Identify what will be transformed into data.**
  Determine which documents, elements, or fields are necessary to support the goals of the project. Focus on information that adds meaningful value for research, access, or community use.
- **Limit information to what is necessary.**
  Apply the principle of data minimization[3] by including only the information needed to meet the project's aims. This includes both structured fields and unstructured content.
- **Exclude nonessential or sensitive information when appropriate.**
  Consider whether certain details introduce unnecessary risk, such as exposing personal information or enabling harmful interpretations. When possible, remove, redact, or limit access to this information.
- **Consider the level of detail (granularity).**
  More detailed data is not always better. In some cases, aggregating information or reducing specificity can help protect individuals and communities while still supporting meaningful analysis.
- **Avoid inferring or assigning attributes.**
  Do not construct attributes such as race, gender, or ethnicity based on appearance, names, or external assumptions. When such information is included, it should be clearly tied to how it appears in the original materials.
- **Evaluate existing categories and data structures.**
  If using or adapting previously created datasets, document who defined the categories and how they were created. Consider whether these categories are appropriate for the current project and its historical and social context.

[3] Definition: The principle that personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. Source: "Art. 5 GDPR – Principles Relating to Processing of Personal Data," n.d.

- **Recognize that categories are historically and socially constructed.**
  Attributes such as race or gender may reflect the perspectives and power structures of the time in which records were created. Treat these categories as contextual rather than fixed or universal.

How to Use This Section

- Use this section when planning what data to extract or create from archival materials.
- Revisit it during data entry or transformation to adjust decisions about inclusion, exclusion, or level of detail.
- Document your reasoning in the Assessment Notes and Actionable Items column, especially when making decisions about sensitive or ambiguous information.

Not all considerations will apply to every project. If a consideration is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Data minimization	Information is intentionally limited to what is necessary to support the project's scholarly, institutional, or community aims; unnecessary data, details, or content are excluded, and the	Information is intentionally limited to what is necessary for the project's aims; unnecessary data or detail is excluded, and decisions are documented.	Some effort is made to limit information to relevant content, but unnecessary data or detail remains; rationale is incomplete or inconsistently documented.	Information is included without clear limits or justification; unnecessary or excessive data is present, and no rationale is documented.	

	rationale for inclusion is documented. Careful evaluation prevents unnecessary collection of sensitive attributes; granularity adjusted when needed to reduce risk of harm or misinterpretation.				

3.2 - Sensitivity and Harm Assessment

This section focuses on identifying information that may require additional care due to privacy, safety, or potential harm. Not all archival materials can or should be made fully open as data. Reviewing materials for sensitivity helps inform decisions about what to include, how to represent it, and how it should be shared.

As you review materials, consider whether they contain:

- **Personally identifiable information (PII)**[4]
  Information that could identify an individual, such as names, addresses, or other direct identifiers.
- **Sensitive demographic information**
  Information related to race, gender, health, sexuality, immigration status, or other characteristics that may expose individuals or groups to risk.
- **Stigmatizing or potentially harmful content**
  Content that could affect reputation, reinforce bias, or contribute to harm if shared without context.
- **Records involving vulnerable populations**
  Materials involving minors, victims, or historically marginalized communities.
- **Language reflecting historical bias or violence**
  Terminology or classifications that reflect unequal power structures or discriminatory practices.

Assessing Potential Harm

In addition to identifying sensitive content, consider how the data might be used and what risks could emerge. This includes:

- Risk of reidentification[5], even in datasets that have been anonymized
- Personal safety risks
- Economic or employment risks
- Legal or residency-related risks
- Social or community harm
- Misinterpretation or misuse of information

This assessment should consider both immediate impacts and downstream uses of the data.

---

[4] Definition: Any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.
Source: McCallister, Erika, Tim Grance, and Karen Scarfone. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). NIST Special Publication (SP) 800-122. National Institute of Standards and Technology, 2010. https://doi.org/10.6028/NIST.SP.800-122.

[5] The process by which anonymized data is linked to outside information to discover the true identity of data subjects, reversing the protections of anonymization. Source: Ohm, Paul. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." SSRN Scholarly Paper No. 1450006. Social Science Research Network, August 13, 2009. https://papers.ssrn.com/abstract=1450006.

Responding to Sensitivity

If risks are identified, possible responses may include:

- Revising what is included in the dataset
- Redacting or removing specific information
- Adding contextual notes or warnings
- Applying access restrictions or tiered access
- Delaying or limiting release

Document the decisions you make and the reasoning behind them.

How to Use This Section

- Use this section during data selection and review to identify sensitive materials.
- Revisit it when making decisions about access, sharing, or publication.
- Record your observations and decisions in the Assessment Notes and Actionable Items column, including any uncertainties or areas requiring further review.

Not all parameters will apply to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
PII evaluation	Comprehensive review of PII; mitigation strategies implemented;	Basic evaluation of PII; appropriate redactions considered.	Incomplete assessment; some exposure risks remain.	No assessment; PII released without review.	

	clear justification for inclusion / exclusion.				
Sensitivity assessment	Comprehensive review of sensitive, harmful, or stigmatizing content; safeguards (e.g., redaction, restriction, contextual notes) are applied.	Sensitive content reviewed and documented; basic safeguards considered or implemented.	Limited assessment; some sensitive materials overlooked or insufficiently addressed.	No documented review for sensitive content; potential risks remain unexamined.	
Risk to individuals or communities	Harm mitigation clearly articulated; includes downstream risks.	Potential harms identified; addressed moderately.	Limited consideration of harm beyond immediate release.	No consideration of harm or misuse.	

3.3 - Indigenous Data and Cultural Knowledge

This section focuses on materials that involve Indigenous communities or culturally specific knowledge systems. These materials may require additional consideration beyond standard privacy or sensitivity assessments.

Some forms of knowledge are not intended for open access or broad distribution, even if they appear in archival records. It is important to recognize that Indigenous communities may have rights, protocols, and expectations regarding how information connected to their communities is described, accessed, shared, and reused.

It is best practice to prioritize Indigenous data sovereignty and stewardship practices by ensuring Indigenous communities lead decision-making processes related to metadata, access, use, and long-term stewardship of the data.[6]

As you review materials, consider whether they include:

- **Culturally restricted knowledge**
  Information that may be governed by community protocols or expectations about who can access or share it.
- **Traditional ecological knowledge (TEK)[7]**
  Knowledge systems related to land, environment, and cultural practices that may be collectively held and context-specific.
- **Ceremonial or sacred information**
  Materials connected to spiritual practices, sacred sites, or ceremonial traditions.

Additional Considerations

- These forms of knowledge may be collectively held, rather than individually owned.
- Access and use may be guided by community-specific protocols, not just institutional policy.
- Open access may not be appropriate, even if the material is not legally restricted.

Where possible, consider:

- Whether consultation with relevant communities is appropriate
- Whether access should be limited, mediated, or restricted
- How to provide context that supports respectful interpretation

[6] Global Indigenous Data Alliance. “Global Indigenous Data Alliance.” Accessed May 21, 2026. https://www.gida-global.org.
and Global Indigenous Data Alliance. “The CARE Principles for Indigenous Data Governance.” Accessed May 21, 2026. https://www.gida-global.org/careprinciples.

[7] The ongoing accumulation of knowledge, practice, and belief about relationships between living beings in a specific ecosystem, acquired by indigenous peoples over hundreds or thousands of years through direct contact with the environment, handed down through generations, and used for life-sustaining ways. Source: “Overview - Indigenous Knowledge and Traditional Ecological Knowledge (U.S. National Park Service).” Accessed May 21, 2026. https://www.nps.gov/subjects/tek/description.htm.

How to Use This Section

- Use this section when working with materials that involve Indigenous communities or culturally specific knowledge systems.
- Revisit it when making decisions about access, description, and reuse.
- Document decisions and uncertainties in the Assessment Notes and Actionable Items column, particularly where community input may be needed.

If this section does not apply to your project, mark it as N/A.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Culturally restricted knowledge	Materials reviewed for culturally restricted knowledge; appropriate consultation, access controls, or contextual safeguards implemented.	Some review conducted; sensitive cultural knowledge identified and partially addressed.	Limited recognition of culturally restricted knowledge; safeguards inconsistent or unclear.	No recognition of culturally restricted knowledge; sensitive materials treated as unrestricted data.	
Traditional ecological knowledge (TEK)	Presence of traditional ecological knowledge identified;	TEK recognized and treated with some caution or contextualization.	TEK present but ethical implications only partially considered.	TEK treated as open data without recognition of cultural stewardship or	

	stewardship practices respect cultural ownership, knowledge sovereignty, and appropriate access limitations.			knowledge sovereignty.	
Ceremonial or sacred information	Ceremonial or sacred materials clearly identified; consultation and appropriate restrictions applied to prevent harm or cultural misuse.	Ceremonial or sacred information recognized and documented; some safeguards applied.	Ceremonial or sacred information inconsistently recognized or insufficiently protected.	Sacred or ceremonial information released without review or consideration of cultural harm.	

4 - How Should This Work Be Represented?

4.1 - Inclusive Description

This section focuses on how materials are described through metadata, descriptive fields, and project language. The words used to describe materials shape how they are discovered, understood, and interpreted.

Description is not neutral. It reflects the perspectives, assumptions, and systems in place at the time it is created. This includes both historical records and the descriptive work done as part of the project.

As you create or revise metadata and descriptive content, consider the following:

- **Review descriptive language for bias.**
  Examine metadata, controlled vocabularies, and project language for terms that may reflect bias, outdated perspectives, or unequal power structures. This may include racialized language, colonial terminology, or historically embedded classifications.
- **Use appropriate and inclusive terminology.**
  Where possible, apply terminology that reflects current understanding and, when relevant, aligns with how individuals or communities identify themselves.
- **Provide context for historical terms.**
  In some cases, original language may need to be retained for accuracy or to reflect the historical record. When this occurs, provide contextual notes to help users understand how and why the language was used.
- **Consider the role of controlled vocabularies.**
  Controlled vocabularies can support consistency and discovery but may also include outdated or limited terms. Review how these vocabularies are applied and whether additional terms or notes are needed.
- **Center people and communities in description.**
  Where appropriate, describe materials in ways that reflect the individuals and communities represented, including language of community origin or preferred language.
- **Align description across the project.**
  Ensure that metadata, descriptive notes, and project-level language (e.g., introductions, documentation, interfaces) are consistent in how they address bias, context, and representation.

How to Use This Section

- Use this section when creating or revising metadata and descriptive content.
- Revisit it during quality review to ensure consistency and clarity across the dataset and project materials.
- Use the Assessment Notes and Actionable Items column to document terminology choices, revisions, and any areas where additional context is needed.

Not all considerations will apply to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Bias-aware description, metadata, and project language	Metadata, descriptive practices, and project language actively address historical bias and power; harmful or outdated terms are contextualized and explained.	Metadata, descriptions, and project language identify and flag harmful or outdated terms; basic context is provided.	Some acknowledgment of bias in metadata, description, or project language, but inconsistently applied or limited in scope.	Bias in metadata, description, and project language is unexamined; harmful, inaccurate, or outdated language remains unaddressed.	

Controlled vocabularies, naming, and preferred language	Inclusive, community-preferred terminology and language used; vocabularies scrutinized.	Vocabularies and language mostly appropriate.	Some outdated or imposed terms remain.	Terms perpetuate harm or misrepresentation.	

4.2 - Context and Provenance

This section focuses on how archival materials are transformed into data while maintaining a clear connection to their original context. Decisions made during transcription, normalization, and structuring can shape how the data is understood and used.

Maintaining context and provenance[8] helps ensure that data remains interpretable, traceable, and reflective of the conditions in which it was created.

As you work with materials, consider the following:

- **Preserve the original context.**
  Consider the historical, social, and institutional conditions in which the records were created. This includes recognizing how power dynamics, language, and recordkeeping practices may have shaped the information.
- **Maintain a clear link to source materials.**
  Ensure that each data point can be traced back to its original document. This allows users to verify information and understand it in context.

[8] The origin or source of something; information regarding the origins, custody, and ownership of an item or collection. (Note: It is a fundamental principle of archives, referring to the individual, family, or organization that created or received the items in a collection. The principle of provenance, or respect des fonds, dictates that records of different origins be kept separate to preserve their context.)
Source: “SAA Dictionary: Provenance.” Accessed May 21, 2026. https://dictionary.archivists.org/entry/provenance.html.

- **Handle uncertainty transparently.**
  When information is unclear, illegible, or ambiguous, indicate this explicitly rather than making assumptions or silently resolving the issue.
- **Approach normalization and standardization carefully.**
  Normalization (making values consistent for comparison) and standardization (applying uniform formats or controlled terms) can improve usability, but may also reduce nuance. Consider how these choices affect meaning, and document how and why they are applied.
- **Preserve meaningful variation.**
  Differences in spelling, naming conventions, or terminology may reflect cultural, historical, or contextual significance. Avoid changing these without careful consideration.
- **Handle demographic data with care.**
  When demographic information is present, record it as it appears in the original materials. Avoid inferring or assigning attributes that are not explicitly stated. Provide context where categories reflect historically specific or biased systems.

How to Use This Section

- Use this section when transcribing, structuring, or transforming materials into data.
- Revisit it during data cleaning or normalization to ensure that decisions do not obscure meaning or introduce inaccuracies.
- Use the Assessment Notes and Actionable Items column to document transformation decisions, especially where interpretation or standardization is involved.

Not all considerations will apply to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Accuracy and faithfulness to original records	Transformation preserves nuance and context; uncertainties documented.	Mostly accurate; basic handling of uncertain data.	Some errors or undocumented assumptions.	Significant distortions; assumptions unacknowledged.	
Normalization and standardization choices	Normalization (making values consistent for comparison) and standardization (applying uniform formats or controlled terms) are clearly defined and justified; decisions balance usability with preservation of historical context.	Normalization and standardization are applied consistently with general justification; some consideration given to preserving meaning and context.	Use of normalization and/or standardization is inconsistent or insufficiently explained; impacts on meaning or context are not fully considered.	Normalization or standardization is applied without justification, obscuring original meaning or introducing historical inaccuracies.	

Handling of demographic data	Avoids inference; records categories transparently; contextualizes historically specific terms.	Demographic practices are mostly responsible.	Some reliance on inference or ambiguous categories.	Inferential or imposed demographic assignments used.	

5 - How Should This Work Be Shared?

5.1 - Access Strategy

This section focuses on how data is made available and who can access it. Access is not a single decision (open vs. closed), but a set of choices that shape how data can be used, interpreted, and reused.

Different types of materials may require different levels of access. In some cases, open access may support research and discovery. In others, limiting access may reduce risk or better reflect the expectations of individuals or communities represented in the data.

As you develop an access approach, consider the following:

- **Identify appropriate levels of access.**
  Determine whether data should be openly accessible, restricted, mediated, limited to specific users or locations, or embargoed for a period of time. In some cases, access may be guided by community review or advisory structures.
- **Consider the sensitivity of the material.**
  Materials that include personal information, sensitive content, or culturally specific knowledge may require additional safeguards or limited access.
- **Account for privacy and expectations of use.**
  Reflect on how individuals or communities represented in the materials might expect the information to be shared or used.
- **Evaluate potential risks of reuse.**
  Consider how data might be used beyond its original context, including the possibility of misinterpretation, misuse, or harm.
- **Consider equity in access.**
  Access decisions can shape who is able to use the data and for what purposes. Consider whether access models create barriers or disproportionately benefit certain groups over others.
- **Incorporate community input where appropriate.**
  When possible, consider how communities or stakeholders may inform decisions about access, including preferences for restriction, mediation, or conditions of use.
- **Use tiered or differentiated access when needed.**
  Not all data needs to be treated the same. Different levels of access can be applied to different parts of a dataset based on sensitivity, risk, or context.

- **Document access decisions.**
  Record how access levels were determined, including any legal, ethical, or community-based considerations. This may include data use agreements or other forms of documentation.

How to Use This Section

- Use this section when planning how data will be shared or published.
- Revisit it when new risks are identified or when access conditions need to be updated.
- Use the Assessment Notes and Actionable Items column to document decisions, including any restrictions, conditions of use, or community input.

Not all access models will apply to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Equity in access	Access strategy intentionally considers equity, balancing openness with the needs and rights of represented communities; barriers to access are minimized while protecting	Access approach generally supports broad access while recognizing some ethical considerations.	Access policies inconsistently address equity or create unnecessary barriers for some users.	Access decisions disregard equity considerations or disproportionately privilege certain users or uses.	

	sensitive materials.				
Community governance or input	Community consulted; shares authority in defining access terms.	Consultation occurred but not fully integrated.	Limited, informal, or symbolic consultation.	No consultation with affected or descendant communities.	
Tiered access models	Access levels (e.g., open, mediated, restricted, onsite, or embargoed) are intentionally designed based on sensitivity, community preferences, and risk assessment.	Some differentiated access controls implemented to address varying levels of sensitivity.	Limited or inconsistent use of access controls; safeguards only partially applied.	No consideration of differentiated access; all materials treated as fully open or unrestricted regardless of risk.	

5.2 - Mitigate Potential Misuse

This section focuses on how data may be used beyond its original purpose. Once data is shared, it may be reused, combined with other datasets, or interpreted in ways that were not anticipated.

Considering potential misuse helps identify risks and provides an opportunity to guide responsible use.

As you prepare data for sharing, consider the following:

- **Identify potential downstream uses and risks.**
  Think about how the data might be used in different contexts, including uses that could result in harm. This may include surveillance, discrimination, profiling[9], or misinterpretation.

  Examples:
  - Government records from regions in conflict could be used to profile individuals, identify activists, or monitor communities.
  - Data on historical property transactions may be misinterpreted as reflecting current ownership, leading to confusion or legal concerns.
- **Consider how context may be lost.**
  Data is often reused outside of its original setting. Without sufficient context, users may draw conclusions that are incomplete or inaccurate.
- **Clearly communicate intended uses and limitations.**
  Describe what the dataset is designed to support and where its boundaries lie. This includes acknowledging gaps, biases, or constraints in the source materials.
- **Provide guidance for responsible use.**
  Include recommendations for citation, interpretation, and appropriate reuse. This may include notes about how to contextualize findings or avoid common misinterpretations.
- **Flag areas that require caution.**
  Identify parts of the dataset that may be particularly sensitive or prone to misuse, and provide additional explanation where needed.
- **Document mitigation strategies.**
  Record any steps taken to reduce risk, such as limiting certain data elements, adding contextual notes, or applying access restrictions.

[9] Definition: Any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements.
Source: GDPR (Article 4). "Art. 4 GDPR – Definitions." General Data Protection Regulation (GDPR), n.d. Accessed May 21, 2026. https://gdpr-info.eu/art-4-gdpr/. Note: Referred to in the GDPR as "Profiling"; the qualifier "Algorithmic" is used here to emphasize the automated, computational nature of the process.

- **Recognize that mitigation is ongoing.**
  Risks may change over time as new technologies, datasets, or uses emerge. Consider how feedback can be received and how updates or corrections can be made after publication.

How to Use This Section

- Use this section when preparing data for sharing or publication.
- Revisit it when new uses or risks become apparent, especially after release.
- Use the Assessment Notes and Actionable Items column to document identified risks, assumptions, and mitigation strategies.

Not all risks will apply to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Assessment of downstream risks	Clear articulation of potential misuse (e.g., surveillance, decontextualized analysis) and mitigation strategies.	General risk assessment included.	Limited discussion of misuse.	No consideration of downstream risks.	
Guidance for secondary users	Explicit use recommendations, limitations, and contextual notes provided.	Some guidance offered.	Minimal guidance; assumptions left to users.	No guidance for responsible reuse.	

6 - How Do We Make This Work Transparent?

6.1 - Documentation of Decisions, Methods, and Limitations

This section focuses on making the project understandable, traceable, and open to review. Clear documentation allows others to understand how the data was created, what decisions were made, and how the data should be interpreted and used.

Transparency supports responsible reuse by helping users assess the reliability, context, and limitations of the dataset.

As you document your work, consider the following:

- **Document how decisions were made.**
  Record the reasoning behind key choices, including what was included or excluded, how data was structured, and how terms or categories were applied. This helps others understand how the dataset was shaped.
- **Describe methods and processes.**
  Provide clear explanations of how the data was created, including transcription methods, data entry processes, normalization or standardization decisions, and any transformations applied.
- **Include documentation that can be shared with users.**
  This may take the form of a readme file, methodology statement, or structured documentation such as a “Datasheet for the Dataset.” Documentation should be accessible to users alongside the data.
- **Acknowledge uncertainty and data quality.**
  Describe known limitations, including transcription challenges, missing data, inconsistencies, or areas where interpretation was required. Where possible, include information about error rates or quality control processes.
- **Clearly state dataset limitations.**
  Explain what the dataset represents and what it does not. This may include gaps in the archival record, uneven representation, or constraints related to the source materials.
- **Maintain provenance and traceability.**
  Ensure that data can be traced back to original sources. Provide citations, identifiers, or links that allow users to locate and review the original materials.
- **Provide ways to verify the data.**
  Include clear pathways for users to check information against source materials or supporting documentation.

- **Document automated tools and processes.**
  If automated methods (e.g., OCR, transcription tools, entity extraction, AI/ML) are used, describe what the tool was used for, how it was configured, and what rules or assumptions shaped the output. Note whether results were reviewed, corrected, or validated before publication.
- **Support review and feedback.**
  Where possible, provide ways for users or communities to offer feedback, identify issues, or suggest corrections.

How to Use This Section

- Use this section throughout the project to record decisions as they are made, rather than waiting until the end.
- Revisit it before publication to ensure documentation is clear, complete, and accessible to users.
- Update documentation over time as new information, corrections, or feedback emerge.

Use the Assessment Notes and Actionable Items column to track what has been documented, where gaps remain, and what may need further clarification.

Not all forms of documentation will apply to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Methodological transparency	Full justification of decisions; Datasheets or similar documentation included.	Reasonable documentation of key decisions.	Partial documentation; gaps in rationale.	Minimal or absent documentation.	

Error rate disclosure	Known error rates, transcription uncertainties, or data quality limitations are clearly documented; sources of error and quality control procedures explained.	Data quality processes described and some acknowledgment of potential errors provided.	Minimal discussion of possible errors; quality control procedures unclear.	No recognition of possible errors or data quality limitations.	
Dataset limitations	Dataset boundaries, scope limits, representational gaps, and potential interpretive constraints clearly documented for users.	Major limitations acknowledged, though documentation may be brief.	Some limitations mentioned but incomplete or unclear.	Dataset presented without acknowledgment of limitations or scope constraints.	
Provenance traceability	All data points traceable to original sources; clear citation pathways.	Most data traceable.	Inconsistent traceability.	Provenance not maintained.	

Verification pathways	Clear mechanisms allow users to verify data through citations, links to source materials, or documentation of original records.	Most data can be traced to original materials through references or documentation.	Verification pathways inconsistent or incomplete.	Data cannot be traced back to original sources or supporting documentation.	

7 - What Are Our Ongoing Responsibilities?

7.1 - Climate Impact and Environmental Sustainability

This section focuses on the environmental impact of creating, storing, and sharing data. Digital projects rely on infrastructure such as servers, storage systems, and computing resources, all of which require energy and materials.

Considering environmental impact helps support responsible stewardship by aligning data practices with long-term sustainability.

As you plan and manage your project, consider the following:

- **Understand the environmental impact of data work.**
  Digitization, data processing, storage, and access systems all require energy and resources. Consider how the scale and design of the project affect these demands.
- **Limit unnecessary data and processing.**
  More data is not always better. Avoid duplicating data, storing information that is not needed, or using resource-intensive processes that do not significantly improve access or understanding.
- **Consider infrastructure and system choices.**
  When possible, select platforms, storage systems, or vendors that demonstrate attention to energy efficiency, responsible data storage, and sustainable digital preservation practices.
- **Plan for long-term stewardship.**
  Data may be stored and maintained over long periods of time. Consider how decisions about retention, storage, and lifecycle management affect environmental impact. This may include decisions about when to retain, update, or deaccession data.
- **Recognize uneven impacts.**
  The environmental effects of digital infrastructure are not evenly distributed. Energy use, resource extraction, and waste associated with data systems may disproportionately affect certain communities.
- **Document sustainability decisions.**
  Record how environmental considerations informed choices about data creation, storage, infrastructure, and long-term management.

How to Use This Section

- Use this section when planning project scope, infrastructure, and workflows.
- Revisit it when making decisions about storage, processing, and long-term stewardship.
- Use the Assessment Notes and Actionable Items column to document considerations, tradeoffs, and any limitations in available options.

Not all considerations will apply equally to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Assessment of environmental impact	Project evaluates environmental impact of data creation, storage, and access; decisions reflect awareness of energy use and infrastructure demands.	Some consideration of environmental impact in project design or infrastructure choices.	Limited acknowledgment of environmental impact; not integrated into decision-making.	No consideration of environmental impact.	
Sustainable data practices	Unnecessary duplication and resource-intensive processes minimized.	Some efforts made to reduce unnecessary data or processing.	Inefficient or excessive data practices present but not fully addressed.	Data practices are excessive, redundant, or resource-intensiv	

				e without justification.	
Infrastructure and vendor sustainability	Infrastructure choices (platforms, storage, vendors) reflect consideration of sustainability, energy efficiency, or responsible digital stewardship.	Some awareness of infrastructure sustainability, though not a primary factor.	Limited consideration of sustainability in infrastructure decisions.	No consideration of sustainability in infrastructure or vendor selection.	
Lifecycle and long-term sustainability planning	Long-term data stewardship includes consideration of environmental impact, retention strategies, and responsible lifecycle management.	Some attention to long-term sustainability and storage implications.	Minimal planning for long-term environmental impact.	No consideration of long-term sustainability or data lifecycle.	

Environmental justice considerations	Acknowledges and reflects on environmental impacts of digital infrastructure on marginalized communities.	General awareness of environmental justice considerations.	Limited or implicit recognition of environmental justice issues.	No recognition of environmental justice considerations.	

7.2 - Iteration, Revision, and Stewardship

This section focuses on how the dataset is maintained and revisited over time. Ethical data practices do not end when a dataset is published. New information, feedback, or uses of the data may require updates or changes.

Planning for iteration and stewardship helps ensure that the dataset remains accurate, responsible, and useful in changing contexts.

As you develop and manage the project, consider the following:

- **Plan for ongoing review.**
  Consider how decisions will be revisited as the project evolves. This may include reviewing data, documentation, or access conditions as new information becomes available.
- **Create pathways for feedback and correction.**
  Provide ways for users, contributors, or communities to report concerns, suggest updates, or request corrections. This may include contact information, feedback forms, or review processes.
- **Update data and documentation as needed.**
  When issues are identified (such as errors, missing context, or new risks) consider how updates will be made and communicated.
- **Consider long-term use and risks.**
  Data may be used in ways that were not originally anticipated. Reflect on how the dataset might be interpreted or reused over time, and how those uses may introduce new risks.

- **Plan for lifecycle management.**
  Consider how long the data will be maintained, how it will be preserved, and under what conditions it may be revised, restricted, or removed.
- **Document changes over time.**
  Keep a record of updates, revisions, and decisions. This helps maintain transparency and allows users to understand how the dataset has evolved.

How to Use This Section

- Use this section when planning for long-term maintenance and stewardship.
- Revisit it after publication to respond to feedback, correct issues, and reassess risks.
- Use the Assessment Notes and Actionable Items column to document review processes, updates, and any changes made over time.

Not all projects will have the same level of ongoing stewardship. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Iterative review processes	Mechanisms for ongoing revision, feedback, and ethical reassessment are built in.	Some review mechanisms present.	Limited responsiveness to new issues.	Static model; no plan for updates.	

Lifecycle risk management	Long-term risks and sustainability addressed (e.g., data permanence, future contexts).	Some attention to long-term concerns.	Minimal consideration of long-term stewardship.	No long-term risk assessment.	

8 - Who Does the Work?

8.1 - Labor Ethics and Acknowledgment

This section focuses on the people involved in creating, transforming, and maintaining the data. Data work is often collaborative and may involve a range of roles, including students, interns, staff, and project partners.

Recognizing how this work is structured and supported helps ensure that contributions are acknowledged and that working conditions are appropriate for the nature of the work.

As you plan and carry out the project, consider the following:

- **Identify who is involved in the work.**
  Consider all contributors across the project lifecycle, including those involved in transcription, data entry, metadata creation, quality review, coordination, and documentation.
- **Recognize different forms of labor.**
  Some contributions may be more visible (e.g., project leadership), while others may be less visible but equally important (e.g., data entry, quality control, or work involving sensitive materials). Consider how these contributions are acknowledged.
- **Consider compensation and working conditions.**
  Reflect on how roles are structured, including workload expectations and compensation. This includes whether essential work relies on unpaid or undercompensated labor.
- **Provide training and support.**
  Ensure that contributors have the information, guidance, and resources needed to perform their work. This is particularly important when working with complex or sensitive materials.
- **Consider opportunities for learning and development.**
  Projects may provide opportunities for skill-building, mentorship, or professional growth, especially for students or early-career contributors.
- **Be aware of roles and decision-making structures.**
  Consider how responsibilities and recognition are distributed across the project. Reflect on whether certain contributions may be less visible or whether decision-making is concentrated in specific roles.

- **Document contributions.**
  Record who contributed to the project and in what capacity. This may include formal attribution methods or project documentation that recognizes different roles.

How to Use This Section

- Use this section when planning project roles and workflows.
- Revisit it during the project to ensure contributors are supported and roles are clearly defined.
- Use the Assessment Notes and Actionable Items column to document roles, responsibilities, and how contributions are acknowledged.

Not all considerations will apply equally to every project. If a parameter is not relevant, mark it as N/A and briefly explain why.

Parameter	Fully Addresses Consideration	Largely Addresses Consideration	Partially Addresses Consideration	Does Not Address Consideration	Assessment Notes and Actionable Items
Non-exploitative labor practices	Project demonstrates fair compensation, appropriate workload expectations, and ethical labor conditions for all contributors; avoids reliance on unpaid or undercompensat	Labor practices generally appropriate; contributors supported, though some gaps may exist.	Evidence of uneven compensation, unclear expectations, or reliance on low-paid labor.	Labor practices are exploitative or rely heavily on unpaid or undervalued work.	

	ed labor for core functions.				
Recognition of invisible and distributed labor	Project explicitly acknowledges all forms of labor, including transcription, data entry, metadata creation, QA, coordination, and interpretive work.	Most forms of labor recognized, though some contributions may be underrepresented.	Limited acknowledgment of labor beyond primary roles.	Labor reduced to a few visible roles; most contributions unrecognized.	
Training, support, and working conditions	Contributors receive appropriate training, supervision, and support; safeguards in place for working with sensitive or potentially harmful materials.	Adequate training and support provided.	Limited training or inconsistent support structures.	No structured training or support; contributors left to navigate work independently.	

Equity and power dynamics in labor structure	Project actively addresses power imbalances; ensures equitable recognition, inclusion, and opportunities for contributors across roles and levels.	General awareness of equity in labor structure.	Limited attention to power dynamics or inequities.	Power imbalances unexamined; credit and authority concentrated without reflection.	

Glossary

Data Minimization

The principle that personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.
Source: “Art. 5 GDPR – Principles Relating to Processing of Personal Data,” n.d.

Extractive Research

A research paradigm rooted in imperial epistemology in which communities are positioned as objects of study rather than as contributors to knowledge. Cultural knowledge and lived experience are treated as raw materials to be collected, classified, and redistributed for the benefit of outside institutions, with little accountability to the communities studied. Contemporary forms persist when research locates the source of a problem within the community itself rather than in the broader structural and historical forces that shaped it.
Source: Smith, Linda Tuhiwai. *Decolonizing Methodologies*. 2nd ed. Bloomsbury Academic, 2012. https://www.bloomsbury.com/us/decolonizing-methodologies-9781786998125/. Chapters 1–4.

Personally Identifiable Information (PII)

Any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual’s identity, such as name, social security number, date and place of birth, mother’s maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.
Source: McCallister, Erika, Tim Grance, and Karen Scarfone. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). NIST Special Publication (SP) 800-122. National Institute of Standards and Technology, 2010. https://doi.org/10.6028/NIST.SP.800-122.

Profiling (Algorithmic Profiling)

Any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements.
Source: GDPR (Article 4). “Art. 4 GDPR – Definitions.” General Data Protection Regulation (GDPR), n.d. Accessed May 21, 2026. https://gdpr-info.eu/art-4-gdpr/.

Note: *Referred to in the GDPR as "Profiling"; the qualifier "Algorithmic" is used here to emphasize the automated, computational nature of the process.*

Provenance

The origin or source of something; information regarding the origins, custody, and ownership of an item or collection. (Note: It is a fundamental principle of archives, referring to the individual, family, or organization that created or received the items in a collection. The principle of provenance, or *respect des fonds*, dictates that records of different origins be kept separate to preserve their context.)
Source: “SAA Dictionary: Provenance.” Accessed May 21, 2026. https://dictionary.archivists.org/entry/provenance.html.

Reidentification

The process by which anonymized data is linked to outside information to discover the true identity of data subjects, reversing the protections of anonymization.
Source: Ohm, Paul. “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization.” SSRN Scholarly Paper No. 1450006. Social Science Research Network, August 13, 2009. https://papers.ssrn.com/abstract=1450006.

Traditional Ecological Knowledge (TEK)

The ongoing accumulation of knowledge, practice, and belief about relationships between living beings in a specific ecosystem, acquired by indigenous peoples over hundreds or thousands of years through direct contact with the environment, handed down through generations, and used for life-sustaining ways.
Source: “Overview - Indigenous Knowledge and Traditional Ecological Knowledge (U.S. National Park Service).” Accessed May 21, 2026. https://www.nps.gov/subjects/tek/description.htm.

Bibliography

Abdilla, Angie, Noelani Arista, Kaipulaumakaniolono Baker, et al. 2020. *Indigenous Protocol and Artificial Intelligence Position Paper*. https://doi.org/10.11573/SPECTRUM.LIBRARY.CONCORDIA.CA.00986506.

Alkemade, Henk, Steven Claeyssens, Giovanni Colavizza, et al. 2023. “Datasheets for Digital Cultural Heritage Datasets.” *Journal of Open Humanities Data* 9 (1). https://doi.org/10.5334/johd.124.

Always Already Computational - Collections as Data. n.d. “The Santa Barbara Statement on Collections as Data.” Accessed May 4, 2021. https://collectionsasdata.github.io/statement/.

"Art. 4 GDPR – Definitions." n.d. *General Data Protection Regulation (GDPR)*. Accessed May 21, 2026. https://gdpr-info.eu/art-4-gdpr/.

"Art. 5 GDPR – Principles Relating to Processing of Personal Data." n.d. *General Data Protection Regulation (GDPR)*. Accessed May 27, 2026. https://gdpr-info.eu/art-5-gdpr/.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. 1st edition. Polity.

Candela, Gustavo, Nele Gabriëls, Sally Chambers, et al. 2023. "A Checklist to Publish Collections as Data in GLAM Institutions." *Global Knowledge, Memory and Communication* ahead-of-print (ahead-of-print). World. https://doi.org/10.1108/GKMC-06-2023-0195.

Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, et al. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19 (November): 43. https://doi.org/10.5334/dsj-2020-043.

Carroll, Stephanie Russo, Desi Rodriguez-Lonebear, and Andrew Martinez. 2019. "Indigenous Data Governance: Strategies from United States Native Nations." *Data Science Journal* 18 (1). https://doi.org/10.5334/dsj-2019-031.

Christen, Kimberly A. 2012. "Does Information Really Want to Be Free? Indigenous Knowledge Systems and the Question of Openness." *International Journal of Communication* 6 (November): 24–24.

Collections as Data: Part to Whole Final Report. n.d. https://doi.org/10.5281/zenodo.10161976.

Cordell, Ryan. 2020. *Machine Learning + Libraries - A Report on the State of the Field*. LC LabsLibrary of Congress. https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf.

Data Lab, Indigenous; Russo Carroll, Stephanie; Suina, Michele; Taitingfong, Riley; Curley, Caleigh; Yracheta, Joe; et al. (2024). Indigenous Data Governance Policy Brief. figshare. Online resource. https://doi.org/10.6084/m9.figshare.26530810.v1

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. The MIT Press.

Dişli, Meltem, Nele Gabriëls, Sally Chambers, Sarah Ames, Beth Knazook, and Gustavo Candela. 2025. "Exploring the Adoption of Collections as Data in the GLAM Context."

Information Research an International Electronic Journal 30 (CoLIS): CoLIS. https://doi.org/10.47989/ir30CoLIS52252.

"Episode 12: Thomas Padilla | Price Lab for Digital Humanities." n.d. Accessed May 4, 2021. https://pricelab.sas.upenn.edu/podcast/2/thomas-padilla/.

Eubanks, Virginia. 2019. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Illustrated edition. Picador.

Farmer, Ashley D., Steven D. Booth, Tracy Drake, et al. 2022. "Toward an Archival Reckoning." *The American Historical Review* 127 (2): 799–829. https://doi.org/10.1093/ahr/rhac161.

Field, Anjalie, Su Lin Blodgett, Zeerak Waseem, and Yulia Tsvetkov. 2021. "A Survey of Race, Racism, and Anti-Racism in NLP." *arXiv:2106.11410 [Cs]*, July 15. http://arxiv.org/abs/2106.11410.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, et al. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92. https://doi.org/10.1145/3458723.

Global Indigenous Data Alliance. n.d.-a. "Global Indigenous Data Alliance." Accessed May 21, 2026. https://www.gida-global.org.

Global Indigenous Data Alliance. n.d.-b. "The CARE Principles for Indigenous Data Governance." Accessed May 21, 2026. https://www.gida-global.org/careprinciples.

Hanna, Alex, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. "Towards a Critical Race Methodology in Algorithmic Fairness." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January 27, 501–12. https://doi.org/10.1145/3351095.3372826.

"Home | CovidBlack." n.d. Accessed May 23, 2022. https://www.covidblack.org/.

Jo, Eun Seo, and Timnit Gebru. 2020. "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (New York, NY, USA), FAT* '20, January 27, 306–16. https://doi.org/10.1145/3351095.3372829.

Mahoney, Jennifer, Roopika Risam, and Hibba Nassereddin. 2020. "Data Fail: Teaching Data Literacy with African Diaspora Digital Humanities." The Journal of Interactive Technology and Pedagogy, December 10.

https://jitp.commons.gc.cuny.edu/data-fail-teaching-data-literacy-with-african-diaspora-digital-humanities/.

McCallister, Erika, Tim Grance, and Karen Scarfone. 2010. *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*. NIST Special Publication (SP) 800-122. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-122.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. Illustrated edition. NYU Press.

Ohm, Paul. 2009. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." SSRN Scholarly Paper No. 1450006. Social Science Research Network, August 13. https://papers.ssrn.com/abstract=1450006.

O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Reprint edition. Crown.

"Overview - Indigenous Knowledge and Traditional Ecological Knowledge (U.S. National Park Service)." n.d. Accessed May 21, 2026. https://www.nps.gov/subjects/tek/description.htm.

Padilla, Thomas. 2025. "Responsible Operations: Data Science, Machine Learning, and AI in Libraries." OCLC, OCLC, June 30. https://doi.org/10.25333/xk7z-9g97.

Padilla, Thomas G. n.d. *Collections as Data: Implications for Enclosure | Padilla | College & Research Libraries News*. https://doi.org/10.5860/crln.79.6.296.

Padilla, Thomas, Hannah Scates Kettler, Stewart Varner, and Yasmeen Shorish. 2023. *Vancouver Statement on Collections as Data*. September 13. https://doi.org/10.5281/zenodo.8342171.

Principles of Indigenous Data Governance. 2020. United States Indigenous Data Sovereignty Network. https://nnigovernance.arizona.edu/sites/default/files/resources/US%2520Indigenous%2520Data%2520Sov%2520Principles%2520Working%2520File.pdf.

"Protocols for Native American Archival Materials." n.d. Accessed March 23, 2026. https://www2.nau.edu/libnap-p/protocols.html.

Rep. Udall, Morris K. [D-AZ-2. 1990. “H.R.5237 - 101st Congress (1989-1990): Native American Graves Protection and Repatriation Act.” Legislation. November 16. 1990-07-10. https://www.congress.gov/bill/101st-congress/house-bill/5237.

Rogers, Jamie, and Rhia Rae. 2025. “Enhancing Access and Research Possibilities through Critical Engagement with Historical Data.” April 2. https://digitalcommons.fiu.edu/record/31525.

“SAA Dictionary: Provenance.” n.d. Accessed May 21, 2026. https://dictionary.archivists.org/entry/provenance.html.

Smith, Linda Tuhiwai. 2012. *Decolonizing Methodologies*. 2nd ed. Bloomsbury Academic. https://www.bloomsbury.com/us/decolonizing-methodologies-9781786998125/.

United Nations Declaration on the Rights of Indigenous Peoples, § Article 31 (2007). https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/11/UNDRIP_E_web.pdf.

University, Santa Clara. n.d. “A Framework for Ethical Decision Making.” Accessed April 1, 2026. https://www.scu.edu/ethics/ethics-resources/a-framework-for-ethical-decision-making/.

“Values Levers: Building Ethics into Design on JSTOR.” n.d. Accessed May 20, 2025. https://www.jstor.org/stable/23474474?searchText=Values+Levers+Building+Ethics+into+Design&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3DValues%2BLevers%253A%2BBuilding%2BEthics%2Binto%2BDesign%26so%3Drel&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&refreqid=fastly-default%3Aa039ae68b196edde627f6ec6b9b31bb0&seq=1.

Veel, Kristin, Catherine D’Ignazio, Annie Ring, Daniela Agostinho, and Nanna Bonde Thylstrup, eds. 2021. *Uncertain Archives* | *Critical Keywords for Big Data*. The MIT Press. https://mitpress.mit.edu/books/uncertain-archives.