# Toward Context-Aware Ethical Data Management

Barbara Oliboni
Department of Computer Science
University of Verona
Verona, Italy
barbara.oliboni@univr.it

Elisa Quintarelli
Department of Computer Science
University of Verona
Verona, Italy
elisa.quintarelli@univr.it

Fabio A. Schreiber
Department of Electronics,
Information and Bioengineering
Politecnico di Milano
Milano, Italy
fabio.schreiber@polimi.it

Kostas Stefanidis
Data Science Research Centre
Tampere University
Tampere, Finland
konstantinos.stefanidis@tuni.fi

Letizia Tanca
Department of Electronics,
Information and Bioengineering
Politecnico di Milano
Milano, Italy
letizia.tanca@polimi.it

## Abstract

Ethical concerns are increasingly central in data management, as data-driven systems support high-stakes decisions. Existing approaches address dimensions such as fairness, transparency, diversity, and data protection, and they do rely on knowledge of the application scenario: which attributes are sensitive, which groups are protected, and which fairness notion applies are typically supplied by a domain expert. That knowledge, however, remains outside the data system. It is fixed once per deployment, applies uniformly to the pipeline it was configured for, and is not itself an object the system can store, validate, reuse, or audit. As a consequence, a single database cannot serve, at the same time, a task in which an attribute must be suppressed and an audit in which the same attribute must be retained. We advocate context-aware ethical data management, in which the mapping from context to ethical requirements is declarative and internal to the system. We introduce a conceptual framework that combines the Context Dimension Tree (CDT) and the Ethical Requirements Tree (ERT), specify when such trees and their configurations are well formed, and define contextual ethical views together with the conditions under which they are correct. Through a personnel management case study, we show how the same data requires different, and sometimes conflicting, transformations across contexts, and we outline the research agenda this opens.

## Keywords

Ethical Data Management, Context-Aware Systems, Data Ethics

## 1 Introduction

Recent advances in data-driven systems and Artificial Intelligence (AI) have significantly amplified the role of data management in high-stakes decision-making processes. From hiring and credit approval to healthcare and public administration, decisions are increasingly supported, or fully automated, by data analytics and machine learning pipelines. As a result, ethical considerations in data management have become a central concern for both researchers and practitioners.

A substantial body of work has addressed key ethical dimensions such as fairness, transparency, diversity, and data protection [15, 24]. These efforts have led to a wide range of techniques for detecting and mitigating bias in datasets, improving explainability, and ensuring compliance with privacy requirements. Each of them is configured once for the deployment it serves: which attributes count as sensitive, which groups are protected, and which notion of fairness applies are decided outside the data system and then held fixed. That configuration, however, cannot be unique: the same attribute or decision criterion may be considered acceptable in one scenario and inappropriate in another one. For example, attributes, such as age or gender, are typically excluded in hiring decisions to avoid discrimination, yet may be explicitly considered in medical or emergency contexts. Similarly, different notions of fairness (e.g., equality vs. equity) may be appropriate depending on the specific goals and constraints of the application.

What is missing is not another fairness, privacy, or transparency metric, but a data management abstraction that determines, at query or pipeline time, which ethical requirements apply, which attributes they affect, and which operations over those attributes are allowed or required. Prior work has studied the ethical dimensions themselves in depth, and context-aware data management has independently studied how context governs data access and analysis; what remains largely unexplored is the interaction between the two, namely contextual reasoning that selects the ethical requirements to be enforced. Where ethical interventions do appear alongside data systems, they appear as external preprocessing steps, governance policies, or auditing procedures. As a consequence, the semantics of ethical adaptation is not represented inside the system, and cannot be reused across tasks, composed with query processing, or audited once the data has been produced. Our goal is to make this context-to-ethics mapping explicit and operational within data management.

To this end, we propose a conceptual framework that jointly models context, through the Context Dimension Tree (CDT) [3, 4], capturing the relevant situational dimensions of data usage, and ethical requirements, through an Ethical Requirements Tree (ERT), representing the possible ethical dimensions, concerns, and conflicts that may arise in the situations captured by the CDT, and guiding the selection of appropriate ethical constraints and transformations.

Our diagnosis is that the contextual knowledge on which ethical data management already depends is held outside the data system, and our contributions make it an object the system represents and acts on, defining a context-to-ethics-to-transformation pipeline:

*EDBT '27, Lille (France)*
© 2026 Copyright held by the owner/author(s). Published on OpenProceedings.org under ISBN 978-3-89318-107-0, series ISSN 2367-2005. Distribution of this paper is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0.

- We introduce the notion of an *ethical context*, a joint CDT-ERT configuration that binds a situational context to a set of ethical requirements, the attributes those requirements affect, and the data management operations they admit, so that requirements are *selected* per task rather than fixed once per deployment.
- We specify when a CDT, an ERT, and their configurations are well formed, which settles what the two tree structures may and may not express, and we define ethical requirements over *sets* of affected attributes with an explicit grouping function, making non-binary sensitive attributes, intersectional requirements, and simultaneously active ethical facets expressible in the model.
- We define *contextual ethical views*, which make ethical adaptation a data management operation between raw data and downstream analytics, together with conditions under which a context-dependent transformation is correct, allowing one database to serve contexts whose requirements are mutually incompatible.
- We derive a research agenda (**RC1**-**RC6**) of problems that arise only once the context-to-ethics mapping is internal to the system: its authoring and evolution, the semantics of ethical views, conflict resolution, query processing, provenance, and benchmarking.

The rest of the paper follows the proposed pipeline. Section 2 identifies the contextual parameters that existing techniques fix in advance, Section 3 specifies how context and ethical requirements are represented, Section 4 defines contextual ethical views and when they are correct, and Section 5 instantiates both on a personnel management database. Section 6 draws the agenda this opens: authoring and evolving the context-to-ethics mapping (**RC1**), semantics and correctness of ethical views (**RC2**), conflicts among simultaneously active requirements (**RC3**), query processing over ethical views (**RC4**), provenance of context-dependent decisions (**RC5**), and benchmarking (**RC6**).

## 2 Background

**Ethical Dimensions in Data Management.** Ethical data management is commonly framed around four core principles [15]. *Fairness* aims to eliminate unwanted bias with respect to protected attributes such as gender, race, or age [31], and is studied both at the individual and at the group level, with the further distinction between equality of treatment and equity of outcome [18]. More recently, there have also been approaches focusing on optimizing fairness across different stakeholders [29, 34]. *Transparency* concerns the ability to interpret how decisions are made, through data provenance [16, 30] and explanations of algorithmic outputs [26]. *Diversity* concerns the representation of heterogeneous entities in a dataset or result set [12], and is often used as a complementary mechanism to promote fairness [21]. *Data protection* concerns the handling of personally identifiable information, which is hardest to guarantee when datasets are combined and apparently anonymised data becomes re-identifiable [33].

These dimensions are frequently in tension: transparency may conflict with privacy, and diversity may be obtained at the expense of fairness or utility. Which of them governs, and how the tension is resolved, varies with the application domain and the situation of use [6, 10, 13, 23, 25], which is the observation this paper builds on.

**Table 1: Ethical data transformations and the contextual parameters they fix before the data is processed.**

| Technique | Operates on | Fixed in advance |
|---|---|---|
| Suppression [19] | Attributes | Which attributes are sensitive |
| Database repair [14, 22, 28] | Tuples/Values | The constraint to be restored |
| Optimized transformation [5, 9] | Distributions | The distortion/utility trade-off |
| Augmentation [27] | Tuples | Which groups are under-represented |
| Massaging [8, 19] | Labels | The target outcome parity |
| Reweighting [19] | Tuple weights | The reference distribution |

**What Existing Approaches Fix in Advance.** A large body of work has focused on preprocessing techniques for mitigating bias in data-driven systems [19]. Representative families are summarised in Table 1: *suppression* removes sensitive attributes and their correlates; *database repair* modifies values to eliminate discriminatory relationships; *optimized transformation* treats fairness as a trade-off against distortion and utility; *augmentation* enriches under-represented groups; *massaging* relabels selected instances; and *reweighting* assigns tuple weights to correct the distribution.

These techniques are not unaware of the application: the entries in the third column are almost always provided as input by a domain expert who knows the scenario, and much of the literature is explicit about this. What is missing is not contextual knowledge, but its status inside the system. That knowledge is supplied at configuration time, holds for the whole dataset or pipeline it was configured for, and is not represented as an artifact the data management system stores, validates, versions, or reasons about. We use the term *context-oblivious* in exactly this sense: the transformation does not take the context as a parameter, even when a human took the context into account when choosing it.

Three consequences follow. First, one database cannot simultaneously support a promotion task in which gender must be suppressed and a fairness audit in which gender must be retained, without duplicating pipelines and re-deciding the configuration out of band for each. Second, since the configuration is external, the choice of requirement is not recorded together with the data it produced, so the result cannot later be audited for *why* a given treatment was applied. Third, nothing prevents two independently configured pipelines over the same base data from being jointly inconsistent, or from jointly revealing what one of them was meant to protect. The closest relatives of our proposal are purpose-based access control and Hippocratic databases [1, 7], which make purpose a first-class notion and enforce it inside the system by rewriting queries against purpose annotated data [20], and responsible data management work that argues for internalising ethical concerns in the pipeline [31, 32]. The difference is the outcome of the decision: those approaches decide *whether* data may be accessed for a purpose, whereas we decide *which representation* of the data is admissible, and produce it.

## 3 Modeling Ethical Contexts: CDT and ERT

This section is concerned with *representation* only: what the two trees are, when they and their configurations are well formed, and who authors them. How a configuration is turned into data is the subject of Section 4.

We propose a framework based on two complementary structures: the Context Dimension Tree (CDT), introduced in [3, 4], and the Ethical Requirements Tree (ERT). The CDT captures the dimensions that characterise a usage context, such as the task at hand, the roles involved, and the organisational setting, while the ERT represents the ethical facets and constraints that may apply. Both trees give a hierarchical and a modular organisation, so that new contextual dimensions or ethical concerns can be incorporated by local refinement rather than by redesign. Their combination yields what we call an *ethical context*, in which ethical requirements are associated with specific situational conditions and thereby determine how data is selected and transformed. Fig. 1 Ⓐ and 1 Ⓑ show an example of each, drawn from a work domain in which employees belong to public or private institutions and hold a job role that may entail risks, and in which financial and family status are also recorded.

**The Context Dimension Tree.** The root of a CDT is drawn as a star-shaped white node. Its children are black nodes representing the context *dimensions*, i.e., the perspectives along which the data relevant to a context is selected, and the children of a dimension are usually white *concept* nodes representing the alternative values that dimension may take. When such values cannot be enumerated explicitly, however, the dimension may instead have an *attribute* node, drawn as a small square, whose domain provides the admissible values. In Fig. 1 Ⓐ, one dimension is the Role of an employee, with concept nodes Clerk, Worker and Manager. When an *attribute* node is attached to a concept node, the attribute parameterises it, as, for example, with the Name of a Public institution.

*Definition 3.1 (Context Dimension Tree).* A *Context Dimension Tree* (CDT), $T$, is a rooted tree whose node set is partitioned into *concept nodes*, *dimension nodes*, and *attribute nodes*, such that: (i) the root is a concept node; (ii) each child of a concept node, if any, is either a dimension node or an attribute node; (iii) the children of a dimension node are either concept nodes or attribute nodes; and (iv) attribute nodes are leaves, each typed with a domain.

A concept node with no dimension children is *atomic*; otherwise it is *refined* by its dimension children.

Conditions (ii) and (iv) separate two roles that are easily conflated. An attribute node does not enumerate alternatives structurally: it provides values from a domain. A dimension introduces alternatives explicitly through concept children, or intensionally through an attribute child, and by (ii) a concept node may itself be refined by further dimensions. The Recruitment node of Fig. 1 Ⓐ is an instance of this recursion rather than an exception to the structure: it is refined by a dimension Contract type with concept nodes Permanent and Fixed-term. Modelling this as a dimension is deliberate, because which of the two applies is itself context-dependent, as contexts $C$ and $C_1$ of Section 5 illustrate.

*Definition 3.2 (Configuration, context).* A *configuration* of a CDT $T$ is a subtree $S$ of $T$ rooted at the root of $T$ such that: (i) for every dimension node in $S$, either its attribute child or exactly one of its concept children belongs to $S$; (ii) for every concept node in $S$, each of its dimension children is either absent from $S$ or present together with its selected child; and (iii) every attribute node in $S$ is either unbound or bound to a value of its domain. A context of a CDT $T$ is a configuration denoted as $C = \{d_1 = c_1(\alpha_1), \ldots, d_k = c_k(\alpha_k)\}$, where each $d_i$ is a selected dimension, $c_i$ its selected concept/attribute name, and $\alpha_i$ is the optional attribute value.

Condition (i) is what makes sibling concept nodes mutually exclusive, since they are alternative values of one perspective, if present in the configuration, whereas sibling dimension nodes are conjunctive: they are orthogonal perspectives that jointly narrow the situation. Condition (ii) permits partial specification, so a context need not fix every dimension. In Fig. 1 Ⓐ, the black node Action specifies the purpose of the analysis, while the remaining dimensions describe the people and organisations involved and their peculiarities. The highlighted nodes form our running context: Institution = Private company (blue), Action = Promotion (yellow), and Role = Clerk (green). The choice of Institution matters and is not decoration: a public institution is subject to different hiring and dismissal regulations, and therefore selects a different requirement set for the same action.

**The Ethical Requirements Tree.** The ERT represents ethical situations in the same way. A path in it identifies an ethical facet, the affected attribute, and, where specified, the constraint or class of admissible transformation; equity with respect to Gender, for instance, triggers a disparity check followed by reweighting or repair. Fig. 1 Ⓑ has two dimensions below the root, the Ethical facet and the Affected attribute. The facets relevant to the situation at hand are Privacy, Fairness, and Diversity, with fairness refined as above. The concept nodes of the affected-attribute dimension are the sensitive attributes of the input data for which guarantees are to be ensured within a given context; in the figure, these are Salary, Race, Gender, Census, and Family situation.

*Definition 3.3 (Ethical Requirements Tree).* An *Ethical Requirements Tree* (ERT) over a schema $\mathcal{R}$ is a Context Dimension Tree in the sense of Definition 3.1 whose root is refined by two mandatory dimensions: Ethical facet, whose concept nodes denote ethical dimensions and may be refined further, and Affected attribute, whose concept nodes denote either an attribute of $\mathcal{R}$ or a *composite* of several attributes.

*Definition 3.4 (Ethical requirement, context-to-ethics mapping).* An *ethical requirement* is a tuple $ER = (F, A, g, \delta, \theta, \mathcal{T})$ in which $F$ is a facet given by a path in the Ethical facet subtree, $A$ is the set of attributes given by the selected Affected attribute node, $g$ is a grouping function mapping tuples over $A$ to groups, $\delta$ is a disparity measure over those groups with threshold $\theta$, and $\mathcal{T}$ is the class of admissible transformations. A *context-to-ethics mapping* $\mu$ assigns to each context $C$ a set $\mu(C)$ of ethical requirements, each obtained from a configuration of the ERT.

Two features of Definition 3.4 settle what the representation can express. Because $A$ is a set and $g$ is explicit, the groups over which a requirement is stated are whatever $g$ induces rather than two by construction: a non-binary sensitive attribute is handled by instantiating $\delta$ with a multi-group measure, such as a max-min ratio of group rates, and an intersectional requirement by letting $g$ range jointly over the attributes of a composite node. Because $\mu(C)$ is a set, several requirements may be active in one context, so fairness and privacy obligations can be represented simultaneously. We obtain this without making ERT dimensions multi-valued, which would alter the semantics of a configuration as established in [3, 4]. Simultaneity, however, is also what makes conflict possible, since two requirements in $\mu(C)$ may prescribe

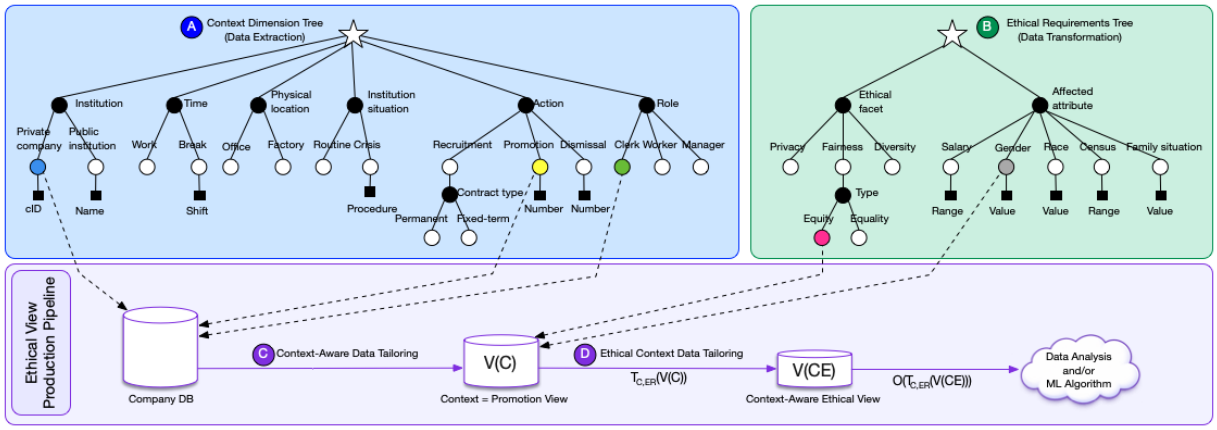



**Figure 1: The CDT Ⓐ and ERT Ⓑ of Section 3, and the Ethical View Production pipeline of Section 4, for a personnel management application. Node shapes: concept (white circle), dimension (black circle), attribute (small square). The highlighted nodes concern the running example: Institution = Private company (blue), Action = Promotion (yellow), and Role = Clerk (green) form the context $C$, while Ethical facet = Fairness/Equity (red) and Affected attribute = Gender (grey) form the requirement set $\mu(C)$ selects. Stage Ⓒ produces the contextual view $V(C)$ and stage Ⓓ the context-aware ethical view $T_{C,ER}(V(C))$, consumed by the downstream operation $O$.**

incompatible treatments of the same attribute; Section 4 states what the resulting transformation must satisfy and Section 5 gives an example.

**Who builds these trees.** The CDT is designed once per application domain by a domain expert together with the database designer, following the CDT design methodology of [4]; it is independent of ethical concerns and can be reused where it already exists. The ERT is authored by the roles that already own compliance, typically a data protection officer, together with a data steward. Its upper levels are largely domain-independent and can be seeded from regulation: special categories of personal data yield the initial values of the Affected attribute dimension, and risk classifications of use cases yield an initial partition of tasks. Its lower levels, and in particular the thresholds $\theta$ and the admissible classes $\mathcal{T}$, are domain-specific. The design effort concentrates on $\mu$, which is authored by humans; our proposal is not to infer it, but to make it an object the system stores, validates, and versions. Acquiring and maintaining it semi-automatically is an open problem (**RC1**).

## 4 Contextual Ethical Views

Given the representation of Section 3, this section defines contextual ethical views, their transformations, and their correctness conditions.

*Definition 4.1 (Ethical context).* An *ethical context* is a pair $(C, ER)$ where $C$ is a context (Definition 3.2) and $ER = \mu(C)$ is the set of ethical requirements the mapping associates with it (Definition 3.4).

An ethical context determines which data is relevant to an operation, which requirements must be enforced over it, which attributes those requirements affect, and which transformations are admissible before the data is used for analytics or learning.

**Contextual views.** Given a schema $\mathcal{R} = \{R_1, \ldots, R_n\}$, each context $C$ selects the relations and attributes relevant to the situation it identifies. This selection is a *contextual view* $V(C)$, a relational expression extracting the data required for the task at hand, and it is the operational dataset associated with that context of use. The data so extracted may still exhibit ethical risks, e.g., group imbalance, discriminatory distributions, disclosure of protected attributes, or lack of representation, so $V(C)$ must be further processed according to $ER$.

*Definition 4.2 (Contextual ethical view).* Let $(C, ER)$ be an ethical context. A *contextual ethical view* is $T_{C,ER}(V(C))$, where $T_{C,ER}$ is a composition of transformations, one for each $ER_i \in ER$, drawn from the admissible class $\mathcal{T}_i$ of that requirement. The dataset consumed by a downstream operation $O$, such as query answering, ranking, statistical analysis, or model training, is $O(T_{C,ER}(V(C)))$.

**Transformation classes.** The admissible classes $\mathcal{T}$ named in a requirement are drawn from the families of Table 1, now parameterised by the requirement rather than fixed in advance. Suppression projects away the attributes in $A$ together with those from which $g$ could be reconstructed. Database repair modifies values so that a constraint derived from $\delta$ and $\theta$ holds. Optimized transformation perturbs the distribution of $V(C)$, trading distortion under $d$ against the disparity $\delta$. Augmentation adds tuples to the groups $g$ identifies as under-represented. Massaging relabels tuples on the outcome attribute so that outcome rates across those groups satisfy $\theta$. Reweighting annotates tuples with weights equalising $\delta$ across groups, leaving values untouched. To these we add annotation, which attaches provenance or explanation metadata rather than altering the data, and which is the class transparency requirements admit; it has no counterpart in Table 1 because the preprocessing literature summarised there addresses bias rather than interpretability. The same family may appear in $\mathcal{T}$ for different requirements with different parameters, which is what it means for the choice to be selected rather than configured.

**When is an ethical view correct?** A transformation that is correct in one context need not be correct in another, so correctness must be stated relative to the ethical context. We require four conditions. *Admissibility*: every operation in $T_{C,ER}$ belongs to the class $\mathcal{T}_i$ of the requirement that motivates it. *Satisfaction*: for each $ER_i \in ER$, the disparity $\delta_i$ measured on the output over the groups induced by $g_i$ does not exceed $\theta_i$. *Minimality*: among the outputs satisfying the previous two conditions, $T_{C,ER}(V(C))$ is at minimal distance from $V(C)$ under a stated distance $d$, in the sense in which repairs of inconsistent databases are minimal [2]. *Non-destructiveness*: base relations are never modified, and every ethical view is derived, which is what allows one database to serve contexts whose requirements are mutually incompatible. Which $d$ makes minimality meaningful for each class, and whether views built for different contexts remain jointly consistent, are open (**RC2**).

Non-destructiveness deserves emphasis because the name REPAIR suggests otherwise. When a requirement admits repair, what is repaired is the contextual view, not the company database: the requirement contributes a constraint over $V(C)$ stated on the groups $g$ induces, and the repair is a minimal modification of the view satisfying it. Base data is left intact, and the repaired view is materialised together with the ethical context that produced it, which is what makes it possible to explain later why an attribute was rebalanced under one context and suppressed under another.

Since $ER$ is a set, $T_{C,ER}$ composes several transformations, and two of them may prescribe incompatible treatments of the same attribute: one requirement suppressing what another needs in order to measure $\delta$. Composition is therefore not always defined, and the resolution is itself part of the ethical context rather than a system default (**RC3**); $C_6$ in Section 5 is an instance.

**System integration.** That ethical concerns should be handled inside the data management pipeline rather than as external post-processing has been argued before [31, 32], and we take it as given rather than as a contribution. What our model adds is the object such an integration needs: a catalog-level representation of $\mu$, with attribute- and view-level annotations recording sensitivity, admissible purposes, applicable requirements, and admissible transformation classes. This lets the system decide not only who may access an attribute (the question answered by access control) but which representation of it is admissible for the task at hand.

## 5 Case Study: Context-Aware Ethical Personnel Management

We instantiate the framework of Sections 3 and 4 on a personnel management database, showing that the ethical treatment of an attribute is context-dependent rather than intrinsic to the attribute itself.

The database covers people, institutions, roles, salaries, performance evaluations, and personal characteristics. The portion of the schema relevant to the running example is the following:

```
PERSON(pID,Surname,Name,DateOfBirth,Gender,Citizenship,
       FamSituation,Census,Race,JobType,. . . )
PREGNANCY(WomanId,IniDate,FinDate)
INSTITUTION(InstName,Type,Domain,Location,CurrentSituation)
EMPLOYEE(InstName,pID,Role,IniDate,Department,Performance)
OFFERS(InstName,JobType)
RISKS(RiskName,JobType)
INCURS(pID,RiskName,GravityLevel)
ROLE-SALARIES(Role,Seniority,JobType,Salary)
```

Attributes such as `Gender`, `Race`, `FamSituation`, or the `PREGNANCY` relation are sensitive in some contexts and irrelevant or inadmissible in others; `PREGNANCY` matters in particular because suppressing `Gender` without it leaves the value inferable.

**The running context.** Let $C$ be the context of Fig. 1: promotion of clerks to managers within a company. The contextual view $V(C)$ joins `EMPLOYEE` and `PERSON` to collect the clerks eligible for promotion, their performance evaluations, and the composition of the current manager population; the downstream operation $O$ may be a top-$k$ query selecting candidates or a model trained to recommend future promotions.

Suppose $\mu(C)$ selects the single requirement

$$ER = (\text{Fairness/Equity}, \{\text{Gender}\}, g, \delta, \theta, \{\text{REPAIR}, \text{REWEIGHTING}\})$$

where $g$ partitions candidates by gender and $\delta$ compares the promotion rates of the resulting groups against their representation among managers. If $\delta$ exceeds $\theta$, e.g., when women are underrepresented among managers relative to the eligible pool, then $V(C)$ does not satisfy $ER$, and $T_{C,ER}$ must produce a view that does.

Concretely, the repair does not modify the company database. The requirement contributes a constraint over $V(C)$ stated on the groups $g$ induces, and $T_{C,ER}(V(C))$ is a minimal modification of the view satisfying it, through reweighting, resampling, or synthetic augmentation, as the governance policy attached to the ethical context permits. The repaired view is materialised together with $(C, ER)$, which is what later makes it answerable why Gender was rebalanced here and suppressed in $C_1$ below.

**Other ethical contexts over the same database.** Other ethical contexts are summarized in Table 2. They show that transformations are context- and requirement-dependent: $C$ and $C_1$ share the same action but select different fairness notions, while $C_4$ and $C_5$ concern race in recruitment but treat it differently for privacy and diversity.

## 6 A Research Agenda for Context-Aware Ethical Data Management

The framework above turns a configuration decision into a stored artifact and, in doing so, raises problems that do not arise when the ethical configuration is external and unique. We state six.

**RC1: Authoring and evolving the context-to-ethics mapping.** $\mu$ is written by humans, so the system must support them: validating well formedness against Definitions 3.1 and 3.3, detecting contexts left unmapped, seeding the ERT semi-automatically from regulatory text and organisational policy, and versioning $\mu$ so that a view produced last year can still be explained under the mapping in force at the time. Regulation and organisational policy change on different schedules from schemas, which is a form of evolution current catalogs do not model.

**RC2: Semantics and correctness of contextual ethical views.** The conditions of Section 4 leave open which distance makes minimality meaningful for each transformation class, and, more importantly, whether views built for *different* contexts over the same base data remain jointly consistent. In our case study, $C_1$ suppresses gender while $C$ retains it in order to measure $\delta$; a user with access to both views must not be able to reconstruct what $C_1$ removed. Cross-context non-interference has not, to our knowledge, been studied for views whose transformations are selected by context.

**Table 2: Ethical contexts over the personnel management database.**

| Context | Action Role | Ethical facet | Affected attribute | Admissible transformation and rationale |
|---|---|---|---|---|
| $C_1$ | Promotion; Clerk | Equality | Gender | Suppression. The gender attribute, together with attributes from which gender may be inferred, is removed from the promotion view. |
| $C_2$ | Promotion; Worker | Equity | Family situation | Reweighting. Candidates with family difficulties receive context-dependent weights that can be used during ranking or candidate selection. |
| $C_3$ | Dismissal; Clerk | Diversity | Gender | Database repair. The decision view is checked for disproportionate dismissal rates across genders and repaired under comparable performance and seniority conditions. |
| $C_4$ | Recruitment; Manager | Privacy | Race | Suppression. The race attribute, together with attributes from which race may be inferred, is removed from the recruitment view. |
| $C_5$ | Recruitment; Manager | Diversity | Race | Augmentation. If some groups are under-represented, the candidate-search or training view is enriched with additional qualified profiles. |
| $C_6$ | Recruitment; Manager | Fairness/Equity and Privacy | Gender and Race | Suppression and reweighting, in conflict. Equity needs group membership to measure disparities; privacy forbids releasing it. The ethical context must resolve the conflict, e.g., via separate auditing and training views. |

**RC3: Conflicts among simultaneously active requirements.** Since $\mu(C)$ is a set, requirements may prescribe incompatible treatments of the same attribute, as in $C_6$. This calls for a means of detecting conflicts statically from the ERT and the admissible classes, and for resolution policies, e.g., priority orders, view decomposition as in $C_6$, or Pareto trade-offs, whose choice is itself part of the ethical context rather than a system default.

**RC4: Query processing over ethical views.** Contexts overlap, so their views do too. This suggests containment and rewriting questions specific to this setting: when can the ethical view for one context be reused for another, which is answering queries using views [17] where the views are indexed by ethical context rather than by query shape, when can an ethical transformation be pushed into a query plan without violating satisfaction, and how should materialised ethical views be maintained incrementally when the context switches, a switch that changes not the data but which transformation is correct.

**RC5: Provenance of context-dependent decisions.** Auditing a context-oblivious pipeline means recording what was done to the data. Here it means recording *why*: the context, the selected requirements, the resolution applied when they conflicted, and the version of $\mu$ in force. Existing provenance models capture the transformation but not the requirement that motivated it.

**RC6: Benchmarking context-dependence.** Current fairness benchmarks largely fix a single context by construction: one dataset, one protected attribute, one notion of fairness. Even where the task is deliberately varied [11], the ethical requirement is not, and therefore cannot measure the property we care about, namely, whether a system selects the appropriate requirement for the situation. A benchmark for this setting needs multiple contexts over shared base data, ground truth for the expected requirement per context, and metrics for cross-context consistency alongside per-context effectiveness.

## 7 Conclusions

We argued that the knowledge which makes ethical data management work in practice, i.e., which attributes are sensitive, which groups are protected, which notion of fairness applies, should be represented inside the data system rather than fixed once per deployment. Combining the Context Dimension Tree with an Ethical Requirements Tree, we specified when such a representation is well formed, defined contextual ethical views and the conditions under which they are correct, and showed on a personnel management database that the same attribute must be suppressed, audited, compensated, or exploited depending on the context. The research agenda opened by this framework lies at the intersection of metadata management, query processing, and provenance.

## Artifacts

This paper does not produce experimental artifacts. It presents a conceptual framework and research vision for context-aware ethical data management; therefore, no datasets, code, or additional materials are associated with this work.

## References


[1] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. 2002. Hippocratic Databases. In *Proceedings of 28th International Conference on Very Large Data Bases, VLDB 2002, Hong Kong, August 20-23, 2002.* Morgan Kaufmann, San Francisco, 143–154. doi:10.1016/B978-155860869-6/50021-4

[2] Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki. 1999. Consistent Query Answers in Inconsistent Databases. In *Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 31 - June 2, 1999, Philadelphia, Pennsylvania, USA.* Association for Computing Machinery, New York, NY, USA, 68–79. doi:10.1145/303976.303983

[3] Cristiana Bolchini, Carlo Curino, Giorgio Orsi, Elisa Quintarelli, Rosalba Rossato, Fabio A. Schreiber, and Letizia Tanca. 2009. And what can context do for data? *Comm. ACM* 52, 11 (2009), 136–140. doi:10.1145/1592761.1592793

[4] Cristiana Bolchini, Carlo Curino, Elisa Quintarelli, Fabio A. Schreiber, and Letizia Tanca. 2009. Context Information for Knowledge Reshaping. *Int. Journal on Web Engineering and Technology* 5, 1 (2009), 88–103. doi:10.1504/IJWET.2009.025015

[5] Rodrigo Borges and Kostas Stefanidis. 2022. Feature-blind fairness in collaborative filtering recommender systems. *Knowl. Inf. Syst.* 64, 4 (2022), 943–962. doi:10.1007/S10115-022-01656-X

[6] Maarten Buyl and Tijl De Bie. 2024. Inherent Limitations of AI Fairness. *Commun. ACM* 67, 2 (2024), 48–55.

[7] Ji-Won Byun and Ninghui Li. 2008. Purpose based access control for privacy protection in relational database systems. *VLDB J.* 17, 4 (2008), 603–619.

[8] Toon Calders, Faisal Kamiran, and Mykola Pechenizkiy. 2009. Building Classifiers with Independency Constraints. In *Proceedings of the 2009 IEEE International Conference on Data Mining Workshops.* IEEE Computer Society, USA, 13–18.

[9] Flávio P. Calmon, Dennis Wei, Bhanukiran Vinzamuri, Karthikeyan Natesan Ramamurthy, and Kush R. Varshney. 2017. Optimized Pre-Processing for Discrimination Prevention. In *Advances in Neural Information Processing Systems.* Curran Associates Inc., Red Hook, NY, USA, 3995–4004.

[10] Gianluca Demartini, Kevin Roitero, and Stefano Mizzaro. 2024. Data Bias Management. *Commun. ACM* 67, 1 (2024), 28–32.

[11] Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. 2021. Retiring Adult: New Datasets for Fair Machine Learning. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual.* Curran Associates Inc., Red Hook, NY, USA, 6478–6490.

[12] Marina Drosou, H. V. Jagadish, Evaggelia Pitoura, and Julia Stoyanovich. 2017. Diversity in big data: A review. *Big data* 5, 2 (2017), 73–84.

[13] Federico Faruffini, Alessandro Corrêa Victorino, and Marie-Hélène Abel. 2021. Vehicle Autonomous Navigation with Context Awareness. In *2nd IEEE International Conference on Human-Machine Systems, ICHMS 2021, Magdeburg, Germany, September 8-10, 2021*. IEEE, Magdeburg, Germany, 1–4.

[14] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and Removing Disparate Impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, NY, USA, 259–268. doi:10.1145/2783258.2783311

[15] Donatella Firmani, Letizia Tanca, and Riccardo Torlone. 2020. Ethical Dimensions for Data Quality. *ACM J. Data Inf. Qual.* 12, 1 (2020), 2:1–2:5. doi:10.1145/3362121

[16] Boris Glavic and Klaus R. Dittrich. 2007. Data Provenance: A Categorization of Existing Approaches. In *DBIS*. Gesellschaft für Informatik e. V., Bonn, 227–241. doi:10.5167/uzh-24450

[17] Alon Y. Halevy. 2001. Answering queries using views: A survey. *VLDB J.* 10, 4 (2001), 270–294.

[18] H. V. Jagadish, Julia Stoyanovich, and Bill Howe. 2022. The Many Facets of Data Equity. *ACM J. Data Inf. Qual.* 14, 4 (2022), 27:1–27:21.

[19] Faisal Kamiran and Toon Calders. 2011. Data preprocessing techniques for classification without discrimination. *Knowl. Inf. Syst.* 33, 1 (2011), 1–33.

[20] Kristen LeFevre, Rakesh Agrawal, Vuk Ercegovac, Raghu Ramakrishnan, Yirong Xu, and David J. DeWitt. 2004. Limiting Disclosure in Hippocratic Databases. In *(e)Proceedings of the Thirtieth International Conference on Very Large Data Bases, VLDB 2004, Toronto, Canada, August 31 - September 3 2004*. VLDB Endowment, USA, 108–119.

[21] Emilia Lenzi and Kostas Stefanidis. 2025. ADAPT: Fairness & diversity for sequential group recommendations. *Inf. Syst.* 133 (2025), 102572. doi:10.1016/J.IS.2025.102572

[22] Kristian Lum and James E. Johndrow. 2016. A statistical framework for fair predictive algorithms. *ArXiv* abs/1610.08077 (2016).

[23] Deborah J. Mugasa, Tom Sorell, Josh Cowls, and Mariarosaria Taddeo. 2024. Ethics in the Governance of Data and Digital Technology: An Analysis of European Data Regulations and Policies. *Digital Society* 3, 11 (2024). doi:10.1007/s44206-024-00101-6

[24] Evaggelia Pitoura, Kostas Stefanidis, and Georgia Koutrika. 2022. Fairness in rankings and recommendations: an overview. *VLDB J.* 31, 3 (2022), 431–458. doi:10.1007/S00778-021-00697-Y

[25] Syama R. and Mala C. 2023. Context-aware behaviour prediction for autonomous driving: a deep learning approach. *Int. J. Pervasive Comput. Commun.* 19, 4 (2023), 477–490.

[26] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of CHI*. ACM, Association for Computing Machinery, New York, NY, USA, 1–13.

[27] Bashir Rastegarpanah, Krishna P. Gummadi, and Mark Crovella. 2019. Fighting Fire with Fire: Using Antidote Data to Improve Polarization and Fairness of Recommender Systems. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, New York, NY, USA, 231–239.

[28] Babak Salimi, Luke Rodriguez, Bill Howe, and Dan Suciu. 2019. Interventional Fairness: Causal Database Repair for Algorithmic Fairness. In *Proceedings of the 2019 International Conference on Management of Data (SIGMOD '19)*. Association for Computing Machinery, New York, NY, USA, 793–810.

[29] Reza Shafiloo, Maria Stratigi, Jaakko Peltonen, Thomas Olsson, and Kostas Stefanidis. 2026. The many facets of fairness in recommender systems: Consumers, providers and items. *Inf. Syst.* 137 (2026), 102643. doi:10.1016/J.IS.2025.102643

[30] Yogesh L Simmhan, Beth Plale, and Dennis Gannon. 2005. A survey of data provenance in e-science. *ACM Sigmod Record* 34, 3 (2005), 31–36.

[31] Julia Stoyanovich, Serge Abiteboul, and Gerome Miklau. 2016. Data Responsibly: Fairness, Neutrality and Transparency in Data Analysis. In *Proceedings of the 19th International Conference on Extending Database Technology, EDBT 2016, Bordeaux, France, March 15-16, 2016, Bordeaux, France, March 15-16, 2016*. 718–719.

[32] Julia Stoyanovich, Bill Howe, and H. V. Jagadish. 2020. Responsible Data Management. *Proc. VLDB Endow.* 13, 12 (2020), 3474–3488.

[33] Anthony Tockar. 2014. Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset. Neustar Research blog. https://agkn.wordpress.com/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/ Accessed 5 Feb. 2026.

[34] Anna Dalla Vecchia, Sara Migliorini, Elisa Quintarelli, and Kostas Stefanidis. 2026. Multi-sided fairness in sequential task assignment. *Inf. Syst.* 139 (2026), 102700. doi:10.1016/J.IS.2026.102700