Refusal, the pipeline, and governance

SIL 503 Week 7 reference packet, 21 October 2026

Module 2 asked whether a classification was biased. This week asks who governs the data at all. Part 1 holds the three cases the opening sort uses, Part 2 the pipeline those cases feed, and Part 3 the governance regimes that answer them. The figures are on the Week 7 page with their sources linked, so cite the original and not this packet.

Put the section number in brackets beside any figure you use, such as [2.1]. The pairs work without laptops for the refusal sort and for the CARE questions on the broker sheet, so everything those two need is printed here.

Part 1

The cases in the sort

1.1 MiDAS, Michigan

Michigan's automated unemployment-fraud system cost $44.4 million to build. It was wrong roughly 93 percent of the time in the fraud determinations it made without human review, falsely accused about 40,000 people, and figures in more than 11,000 bankruptcy filings.

The class action, Bauserman v. Unemployment Insurance Agency, was allowed to proceed by the Michigan Supreme Court in 2022. No one scored by the system had a way to refuse the scoring, which is the criterion the opening sort applies.

Source: Angwin (2022), The Markup; Charette (2018), IEEE Spectrum, on the cost of the system.

1.2 SafeRent tenant screening

SafeRent paid $2.28 million in 2024 in a Fair Housing Act settlement over an AI tenant-screening score that penalized Black and Hispanic applicants using housing vouchers.

An applicant is scored by a screening company the applicant never chose and cannot decline, because declining the score means declining the apartment.

Source: Cohen Milstein, on the $2.28 million settlement agreement.

1.3 The Allegheny Family Screening Tool

Carnegie Mellon researchers found the tool flagged 32.5 percent of Black children for mandatory neglect investigation against 20.8 percent of white children. Caseworkers disagreed with the algorithm's score about a third of the time, and a technical glitch fed workers wrong risk scores for more than two years.

The families scored had no say in the decision to score them. Week 3 worked this case as a bias problem; here it is a governance problem, and the numbers are the same ones.

Source: WESA (2022), on the Allegheny Family Screening Tool.

Part 2

The surveillance pipeline

2.1 American Dragnet

Georgetown's investigation found that ICE has run face-recognition searches on the driver's-license photos of roughly one in three US adults, can access license data on three in four, and tracks vehicle

movements in areas covering about 70 percent of the adult population. Utility hookup records can locate a mover's new home address.

Face recognition ran on Rhode Island DMV records by 2008, a decade before the practice became public. Georgetown puts ICE spending on surveillance and data infrastructure at about $2.8 billion between 2008 and 2021.

Source: Wang, McDonald, Bateyko and Tucker (2022, updated 2025), American Dragnet, Georgetown Center on Privacy and Technology.

2.2 The December 2025 awards

In December 2025 ICE awarded 13 private firms skip-tracing contracts under a vehicle worth up to $1.2 billion over two years, with contractors receiving as many as 50,000 names a month to locate using AI tools.

Two separate figures

The $1.2 billion describes one 2025 contract vehicle. It is additional to, and distinct from, the $2.8 billion of cumulative spending at 2.1. A memo that adds them together has made an error a reader will find.

Source: American Immigration Council (2026), on ICE's private bounty hunters and AI skip-tracing.

2.3 The withdrawn broker rule

The federal rule that would have treated data brokers as consumer reporting agencies was withdrawn on 15 May 2025.

EPIC documents ICE and CBP buying mobile-location, license-plate and jail-booking data from brokers, including in places where local policy prohibits sharing information with ICE directly. A purchase does what a subpoena could not, and no court reviews it, so a local rule against sharing closes one door in a building with several.

Source: EPIC (2025), on the CFPB's withdrawal of the proposed data-broker rules; EPIC (2025), How data brokers harm immigrants.

Part 3

Governance in practice

3.1 Te Hiku Media and the Kaitiakitanga licence

Te Hiku Media, a Maori community media organization in Aotearoa New Zealand, ran the Korero Maori campaign and gathered about 300 hours of community-donated audio in ten days. The resulting recognizer has an accuracy of roughly 92 percent for te reo Maori speech and 82 percent for bilingual speech, which gives the language a tool most minority languages still lack.

Te Hiku released the work under a Kaitiakitanga, or guardianship, licence, restricting use to the benefit of Maori, and refused to sell when large technology companies asked. The licence's operating rule is that any benefit derived from the data flows back to its source, and the community that donated its voices keeps authority over what those voices train.

Source: Winkless (2026), IEEE Spectrum; Te Hiku Media, Papa Reo; the Kaitiakitanga licence text, published on GitHub.

3.2 Negotiated release at Waikato

At the University of Waikato, Te Taka Keegan's team built a te reo Maori text-to-speech voice from 7 hours and 45 minutes of recordings donated by a single speaker, Ngaringi Katipa, with a 6.78 percent word error rate.

Instead of publishing the voice openly, Keegan is negotiating release terms with the three iwi affiliated with the donor: Waikato, Maniapoto and Raukawa. Open release is the default in machine learning, and this is a deliberate departure from it.

Source: Winkless (2026), IEEE Spectrum, on Maori data sovereignty and AI voice models.

3.3 The CARE principles

Collective Benefit, Authority to Control, Responsibility, Ethics. CARE puts people and purpose alongside the data-centric FAIR principles, and the drafters' own summary of the pairing is “Be FAIR and CARE.”

The principles were drafted on 8 November 2018 at an International Data Week workshop in Gaborone, Botswana, by thirteen drafters from seven countries, building on the Maori data-sovereignty network Te Mana Raraunga, the US Indigenous Data Sovereignty Network, and the Aboriginal and Torres Strait Islander collective Maiam nayri Wingara, whose own five principles are control, contextuality, relevance, accountability and protection.

Source: Carroll et al. (2020), The CARE principles for Indigenous data governance, Data Science Journal, doi 10.5334/dsj-2020-043.

3.4 OCAP, in force since 1998

Ownership, Control, Access, Possession. OCAP is older than CARE and further into practice: it has been in force since 1998 and governs the First Nations Regional Health Survey, backed by community data-sharing agreements.

A 2025 five-phase framework describes the road from no governance to fully Indigenous-led governance in practice. Scholars also warn that community consent does not substitute for individual consent, so the two are not alternatives.

OCAP beyond the IRB standard

A university review board requires individual informed consent. OCAP-style governance also requires collective consent, ongoing community control, and the right to withdraw the entire dataset. No consent form in this building offers the third.

Source: First Nations Information Governance Centre, the First Nations principles of OCAP; Prehn (2025), Australian Journal of Social Issues, doi 10.1002/ajs4.70086; McNealy (2024), CHIRON.

3.5 Three US tribal examples

The Pueblo of Laguna partnered with the University of New Mexico to build its own census software while keeping ownership of the resulting data. The Swinomish Tribe's climate-change data agreements require tribal leadership approval before any external use or publication. The Navajo Nation's human research review board has required researchers to transfer their data to the Nation at project completion since 1996.

Each is a contract term somebody negotiated, and each is copyable into an agreement a small organization could sign.

Source: Carroll, Rodriguez-Lonebear and Martinez (2019), Indigenous data governance: strategies from United States Native nations, Data Science Journal.

3.6 The Rohingya registration

After UNHCR's 2018 biometric registration of Rohingya refugees, Human Rights Watch found that roughly 830,000 names with biometric data went to Myanmar, the government the refugees had fled, for repatriation-eligibility screening.

Of 24 refugees Human Rights Watch interviewed, one recalled being asked about sharing data with Myanmar. Most had received English-only consent receipts with the yes boxes already checked, and 21 people whose names appeared on verified repatriation lists went into hiding out of fear of forced return.

UNHCR said publicly that the registration data was not linked to repatriation while using it for that purpose, and it never completed a data protection impact assessment. A registration built to deliver aid became, through one transfer, a screening list held by the government the refugees were fleeing.

Source: Human Rights Watch (2021), UN shared Rohingya data without informed consent.

3.7 The organization in the pitch

Not in the packet

Nothing published here says what the organizations in this room hold, who governs it, or what any of them agreed to when they collected it. The CARE questions on the broker sheet ask who authorized a collection and who could refuse it now, and for a real organization those answers come from a person.

Write down whom you would ask, and which document would answer it: a data-sharing agreement, an intake consent form, a funder reporting requirement, or a vendor contract. That list is the start of the practice each person names at the close.

Source: The Week 7 page, on CARE and OCAP as working governance.