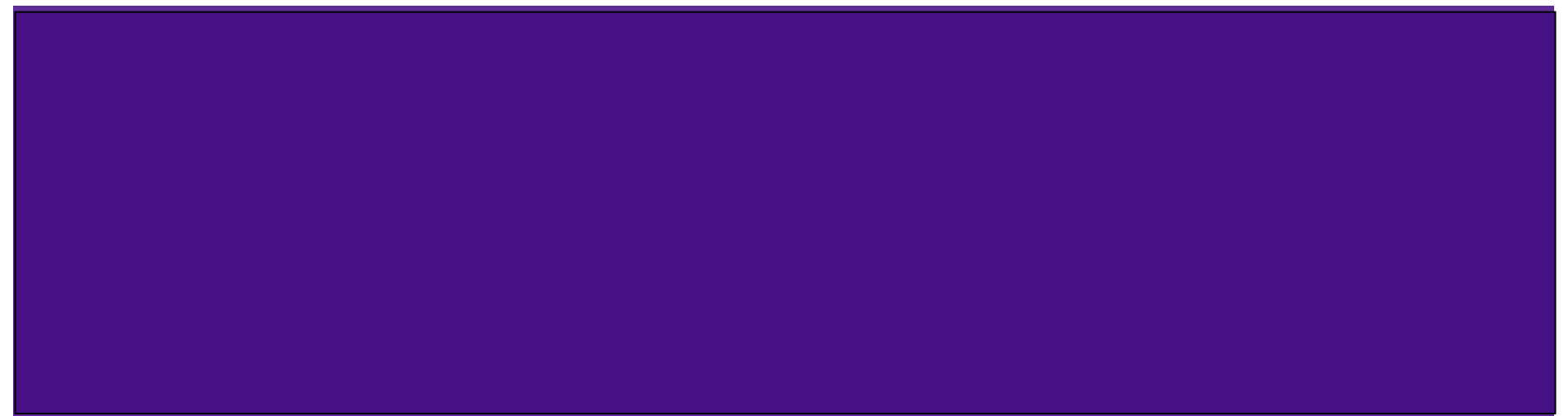# CSDE 502
Data, Pt. I

April 28, 2023

Population Data Systems (Seltzer and Anderson, 2001)

- One-time comprehensive data-gathering, e.g. censuses
- One-time or periodic data-gathering via sample, e.g. DHS, ACS, sample registries
- Comprehensive administrative-reporting systems, e.g. national population registries

Types of Population Data (Seltzer and Anderson, 2001)

- **Microdata** - individuals are identifiable through unique IDs or names & addresses + membership in a vulnerable group
- **Mesodata** - aggregated microdata, but some levels are very small geographic areas or groups
- **Macrodata** - aggregated microdata to large geographic areas or groups
- **Other** - everything that goes into collecting & disseminating population data

Colonial Origins of Population Data Systems (S & A, 2001)

- **The Domesday Book -** compiled by William the Conqueror in 1085; he needed taxes
- **The US Census and Special Censuses -** the 1870 census was the first to enumerate Indigenous peoples in what is now the U.S.
  - Francis Walker made the change to distinguish between the "constitutional" and "true" population
  - Treaties with Nations codified "special censuses" as early as 1817 via Bureau of Indian Affairs
  - Non-treaty censuses after the Indian Removal Act of 1830 ro the Trail of Tears in the Cherokee, Choctaw, and Creek Nations

Colonial Origins of Population Data Systems (S & A, 2001)

- **Rwanda Population Registration System**
  - Created by Belgian colonizers in 1930s, led to more fixed concept of ethnic groups: Hutu and Tutsi
    - At the time of creation, the colonial state was pro-Tutsi for "pseudoscientific racial grounds"
    - Around the time of Rwandan independence, Belgians began to support Hutus instead
  - Maintained until the start of the Rwandan genocide of the Tutsi people
- **Countless other African censuses and registries in apartheid nations**

Principles of Data Governance (Carroll et al, 2020)

DATA PRINCIPLES						
INDIGENOUS				MAINSTREAM		
New Zealand Indigenous Data Sovereignty Principles	Australia Indigenous Data Sovereignty Protocols	United States Indigenous Data Governance Principles	Canada Indigenous Data Governance Principles	Open Data Charter Principles	FAIR Principles for Data Management and Stewardship	STREAM Properties for Industrial and Commoditized Data
Authority	Self-Determination	Inherent Sovereignty	OCAP®	Open By Default	Findable	Sovereign
Relationships	Available and Accessible	Indigenous Knowledge	Indigenous Knowledge	Timely and Comprehensive	Accessible	Trusted
Obligations	Collective Rights and Interests	Ethics	Methodology and Approaches	Accessible and Usable	Interoperable	Reusable
Collective Benefit	Accountability	Intergenerational Collective Wellbeing	Evidence to Build Policy	Comparable and Interoperable	Reusable	Exchangeable
Reciprocity	Exercise Control	Relationships	Ethical Relationships	For Improved Governance & Citizen Engagement		Actionable
Guardianship			Data Governance	For Inclusive Development and Innovation		Measurable

People oriented principles	Purpose oriented principles	Data oriented principles

Principles of Data Governance (Carroll et al, 2020)

CARE Principles and Subprinciples

1. **Collective Benefit**
   a. Inclusive development & innovation
   b. Improved governance & citizen engagement
   c. Equitable outcomes
2. **Authority to Control**
   a. Recognizing rights and interests
   b. Data for governance
   c. Governance of data
3. **Responsibility**
   a. Positive Relationships
   b. Expanding capability & capacity
   c. Indigenous languages & worldviews
4. **Ethics**
   a. Minimizing harms & maximizing benefits
   b. Justice
   c. Future use

FAIR Principles

1. **Findable**
2. **Accessible**
3. **Interoperable**
4. **Reusable**

CARE in Practice (Carroll et al, 2021 - *Nature*)

Practice 'CARE' in data collection	Engage 'CARE' in data stewardship	Implement 'CARE' in data community	Use 'FAIR' with 'CARE' in data applications
Define cultural metadata	Use appropriate governance models	Indigenous ethics inform access	Fairness, Accountability, Transparency
Record provenance in metadata	Make data 'FAIR'	Use tools for transparency, integrity and provenance	Assess equity

Safeguards (Seltzer and Anderson, 2001)

- **Don't collect it**
- **Use samples instead of population**
- **Induce errors**
- **Use decentralized storage or linkage data**
- **Legal protections**
- **Creation and updating of ethical systems**