## Library Example

The library promises it doesn't keep members' reading history.

```
CREATE TABLE LOANS (
    loan_id        INT  NOT NULL AUTO_INCREMENT,
    book_id        INT  NOT NULL,
    member_id      INT  NOT NULL,
    checked_out_on DATE NOT NULL,
    due_on         DATE NOT NULL,
    returned_on    DATE NULL,
    PRIMARY KEY (loan_id),
    FOREIGN KEY (book_id)   REFERENCES BOOKS (book_id)     ON DELETE RESTRICT,
    FOREIGN KEY (member_id) REFERENCES MEMBERS (member_id) ON DELETE CASCADE
);
```

- every returned loan stays forever, along with its member_id

- options
  - clear member_id once a loan is returned
  - delete loans after a retention period
- costs and wrinkles
  - fines and disputes need recent history
  - ON DELETE RESTRICT on books blocks deleting any book that has ever been loaned – if loan history is cleared, books not borrowed recently can be deleted
  - the audit trail records member ids



## Stakeholders

- a loan record isn't simply "the library's data" – many parties have a claim on it
  - the member it describes – it is about them and reveals what they read
  - the library that collected it – it needs the record to run lending and fines
  - vendors that receive it – an e-book platform sees what members read and may want to use that
  - law enforcement requests – may request records in an investigation
  - society – libraries matter as places to read without being watched
- these claims compete – no single party owns the data outright
  - *stakeholders* are everyone with a claim on the data or who is affected by how it is used



## Stakeholder Analysis

- identify the stakeholders
- for each, identify what they gain and what they risk

| stakeholder | gains | risks |
|---|---|---|
| members | books, reading recommendations | reading history exposed or misused |
| library | operations, usage statistics | breach, loss of trust, legal liability |
| e-book vendor | usage data, product improvement | regulation, reputation |
| researchers | data on reading habits | re-identifying people by mistake |

- where do the interests conflict?
  - library's stats and the vendor's usage data vs members' expectation that reading is private



## Managing Conflicts of Interest – Step 1

- look for designs that ease the conflict
  - e.g. the library's stats need counts, not reading histories – keep per-book loan counts instead of per-member history
  - e.g. the e-book vendor gets only what it needs, under a contract that limits use and retention
  - e.g. members can opt in to keeping their reading history

## Managing Conflicts of Interest – Step 2

- reason through what remains using ethical frameworks
  - questions to consider
    - whose harm is worse and harder to undo?
    - what were people told to expect?
    - who has less power or choice?
      - e.g. members can't negotiate with the library
    - what does the law require, as a floor rather than a goal?
  - lenses
    - consequences – which option cause the least harm overall?
    - rights and duties – what is owed to people, whatever the benefit?
    - fairness – who bears the costs and who gets the benefits?
    - character – what would a responsible professional do?
  - the judgment needs the most care in cases where the lenses disagree



## Managing Conflicts of Interest – Step 3

- explain and document the decision
  - say which considerations decided it
  - tell the people affected
  - revisit as things change



## Custodianship Across the Data Life Cycle

- the organization holding the data is responsible for the data at every stage, not only while it is in active use
  - also applies to the team that runs the databases

| stage | custodian's responsibility |
|---|---|
| creation | which columns exist, and why |
| processing | access control, transactions, validation |
| review and reporting | views, reports, inference risks |
| retention and retrieval | backups, archives, retention schedules |
| destruction | deletion everywhere: tables, backups, logs, copies |



## Professional Responsibility

- professional codes of ethics describe the responsibilities individuals take on by working in a field
  - to the public, to the people their work affects, and to clients and employers
- ACM Code of Ethics and Professional Conduct (https://www.acm.org/code-of-ethics)
  - 1.6 – respect privacy: use personal data only for legitimate purpose, collect only what's needed, and set clear retention and disposal policies
  - 1.7 – honor confidentiality: protect information entrusted to you
  - 2.9 – design and implement systems that are robustly and usably secure
- applies to the individuals who design and build systems, not only to the organizations that run them
  - due care is personal too