Lesson 8.1 Exercises: Digital Identity and the Data Economy

Module 8: The Future Problem

Prof. Dr. Joerg Osterrieder

Digital Finance — BSc Course

(c) Joerg Osterrieder 2025–2026

Exercise 1: Identity Model Comparison

Scenario: A small country is designing a national digital identity system for its 5 million citizens. The system must support bank account opening, government services, and mobile phone registration.

Tasks:

a. Describe how each of the three identity models (centralized, federated, self-sovereign) would work in this context. For each, explain who issues the identity, where it is stored, and how it is verified.

b. Create a comparison table with the columns: User control, Single point of failure risk, Implementation cost, and Time to deploy. Rate each model as High / Medium / Low for each attribute.

c. The country has limited technical infrastructure. Which model do you recommend for Phase 1 deployment, and why?

d. How might the country **migrate** from your Phase 1 recommendation toward self-sovereign identity in Phase 2?

Difficulty: Introductory–Intermediate — tests understanding and comparative analysis.

(c) Joerg Osterrieder 2025–2026

Exercise 2: Designing a Verifiable Credential

Scenario: A university wants to issue digital diplomas as verifiable credentials that graduates can use to prove their qualifications to employers and banks (for student loan verification).

Tasks:

a) List the **claims** (data fields) that should be included in the credential. Distinguish between mandatory fields (needed for verification) and optional fields (nice to have).

b) Draw the "trust triangle" for this use case, labeling the Issuer, Holder, and Verifier. Include at least two different verifier scenarios.

c) Explain how a bank could verify the credential **without contacting the university**.

d) A graduate loses their phone (digital wallet). Describe a recovery mechanism that maintains the security properties of the credential.

Difficulty: Intermediate — requires applied design thinking.

(c) Joerg Osterrieder 2025–2026

Exercise 3: Zero-Knowledge Proof Scenarios

Scenario: A FinTech lender wants to use zero-knowledge proofs to streamline loan applications while maximizing customer privacy.

Tasks:

a. For each of the following loan requirements, describe what a ZKP would prove and what remains private:
   - Minimum income threshold ($60,000/year)
   - No bankruptcy in the past 7 years
   - Residency in a specific country
   - Age ≥ 21

b. Identify one loan requirement where a ZKP is **not sufficient** and the lender genuinely needs the underlying data. Explain why.

c. A competitor argues that ZKPs are "security theater" because the lender still needs to trust the credential issuer. Evaluate this argument: is the competitor correct? Why or why not?

Difficulty: Intermediate — requires conceptual reasoning about cryptographic properties.

Exercise 4: Valuing Personal Financial Data

Scenario: A data broker sells anonymized consumer financial profiles. Each profile contains transaction history, estimated income bracket, and spending category breakdown. The broker has 10 million profiles.

Given (all synthetic):

- Hedge funds pay \$0.15 per profile per month for transaction data feeds
- Advertisers pay \$0.05 per profile per month for spending category data
- Credit bureaus pay \$0.08 per profile per month for income estimates
- Operating costs: \$500,000 per month (servers, compliance, staff)

Tasks:

a. Calculate the broker's **monthly gross revenue** from all three buyer segments.

b. Calculate the **monthly profit** (revenue minus operating costs).

c. Calculate the **annual revenue per consumer profile**. Compare this to the value the consumer receives (typically \$0).

d. If regulators required data brokers to pay consumers 30% of revenue derived from their data, what would each consumer receive per year? Would this change consumer behavior?

Difficulty: Intermediate — requires arithmetic and economic reasoning.

Exercise 5: Choosing a Privacy-Preserving Technique

Scenario: You are a consultant advising three different financial institutions. Each has a different privacy challenge:

Client	Need	Constraints
A (Central bank)	Publish aggregate economic statistics without exposing individual bank data	Must be mathematically provable; some accuracy loss acceptable
B (Insurance consortium)	Jointly model risk across 4 insurers without sharing policyholder data	Must produce exact results; moderate performance OK
C (Cloud FinTech)	Run credit scoring on customer data hosted in the cloud	Must be fast; trusts the cloud provider's hardware

Tasks:

a. For each client, recommend the **most appropriate** privacy-preserving technique and justify your choice.

b. For each client, explain why at least one **alternative technique** would be a poor fit.

c. Client B's CEO asks: "Why can't we just anonymize the data and share it?" Write a two-paragraph response explaining the risks of relying solely on anonymization.

Difficulty: Advanced — requires matching techniques to real-world constraints.

Exercise 6: Alternative Data and Ethical Boundaries

Scenario: A hedge fund considers purchasing three alternative data feeds to predict retail stock performance:

1. **Satellite parking lot data:** Drone imagery counts cars at 500 retail locations weekly
2. **Anonymized credit card transactions:** Aggregated spending at major retailers, sourced from a payment processor
3. **Employee sentiment:** Scraped from anonymous workplace review websites

Tasks:

a) For each data source, assess whether the individuals whose data is used (shoppers, cardholders, employees) **consented** to this specific use. Use a scale: Explicit consent / Implied consent / No consent.

b) For each data source, identify the potential for **harm to individuals** if the data is misused or re-identified.

c) Draft a 3-point "ethical data use policy" that the hedge fund should adopt. Each point should be specific and actionable.

d) A regulator proposes requiring all alternative data providers to obtain explicit consent from data subjects. Evaluate: would this policy be **effective**, or would it effectively destroy the alternative data market?

Difficulty: Advanced — requires ethical reasoning and policy analysis.

Exercise 7: Comprehensive Case – Portable Financial Identity

Scenario: A migrant worker moves from Country A to Country B. In Country A, she has:

- A verified bank account (3 years of history)
- A good repayment record on a microloan
- No formal credit score (Country A lacks a credit bureau)

In Country B, she needs to: (1) open a bank account, (2) rent an apartment (requires creditworthiness proof), and (3) send remittances home.

Tasks:

a. Under the **current system** (centralized identity, no cross-border portability), describe the barriers she faces for each of the three needs.

b. Design an **SSI-based solution** using verifiable credentials and ZKPs. For each need, specify: what credential is presented, what is proved, and what stays private.

c. Identify **two realistic barriers** to implementing your SSI solution in practice.

d. Calculate the cost savings if SSI-based KYC costs \$2 per verification vs. \$25 for traditional cross-border KYC. Assume 50 million cross-border workers globally, each requiring 3 verifications per year.

Difficulty: Advanced–Integrative — combines identity, privacy, and financial inclusion.

Exercise 8: Designing a Privacy-First Data Marketplace

Scenario: You are tasked with designing a financial data marketplace where individuals can **sell access to their own financial data** to banks, insurers, and researchers — while maintaining control over what is shared.

Tasks:

a. Design the marketplace architecture. Draw a diagram showing the roles of: data owner (individual), data buyer (institution), marketplace platform, and consent manager. Show how data and payments flow.

b. For each of the following buyer requests, specify which privacy technique you would use and what the buyer receives:
   - A bank wants to check if a customer's income exceeds $40K
   - A researcher wants aggregate spending patterns across 10,000 users
   - An insurer wants to run a risk model on a customer's health spending

c. Propose a **pricing model** for the marketplace. How should the data owner be compensated? Should pricing be per-query, subscription, or auction-based?

d. Identify the **two biggest risks** that could cause this marketplace to fail, and propose a mitigation for each.

Difficulty: Advanced–Creative — requires system design and business model thinking.