

A Scottish Databanking Pilot

A proposal for a regulated personal-data intermediary demonstrator

Dr Dan Protopopescu, University of Glasgow, Scotland, UK | databanking.org | September 2026

— Concept note

Databanking is a proposed model for regulated personal-data intermediation in which personal data remains under the custodial control of trusted intermediaries, while approved computations are brought to the data and only authorised, policy-constrained outputs are released.

The central insight is simple: organisations frequently do not need a copy of a person's underlying data. They need an answer derived from it.

A service may need to know whether someone is over 18, professionally licensed, financially eligible, resident at an address, entitled to a service, or satisfies a defined criterion. Under prevailing data architectures, answering such questions commonly requires the individual to disclose records from which the organisation derives the answer itself. The organisation consequently acquires, stores and assumes responsibility for substantially more personal information than may be necessary for the transaction.

Databanking reverses this relationship: the data remains with a regulated custodian, the organisation submits an authorised query or computation, and the Databank returns only the permitted result.

The model is described in the Databanking white paper: <https://doi.org/10.2139/ssrn.6992518>, with further material at: <https://databanking.org>

This proposal suggests establishing a **limited Scottish Databanking pilot** to determine whether the model is technically, institutionally and commercially viable in a carefully selected sector.

The pilot would not attempt to create a universal repository for all personal data. Instead, it would test Databanking first as a **specialised, regulated infrastructure model for a narrow class of high-value transactions**.

— 1. The proposition

The pilot would test whether a neutral personal-data intermediary can allow organisations to obtain trustworthy answers about individuals without acquiring the underlying personal records from which those answers are derived.

For example, rather than an organisation requesting documentation and independently determining whether a condition is satisfied, it could ask a Databank:

Does this person satisfy criterion X?

The Databank would determine the answer from authoritative data under its custody and return an appropriately authenticated result.

Depending on the use case, outputs might include attestations, cryptographic tokens, match/no-match results, eligibility decisions, aggregates, aliases or other strictly constrained information.

The underlying records would remain within the Databank environment.

This separation between **data custody and data analysis** is fundamental to the model. A Databank would not be a conventional cloud provider or data broker. Its economic and legal incentives would need to be aligned with its custodial role. It should earn revenue from services such as custody, authorised queries and computation rather than from exploiting, selling or profiling the information entrusted to it.

The longer-term proposition is therefore not merely a new software service, but potentially a new category of **regulated digital infrastructure institution**.

— 2. Why begin with a pilot?

The Databanking architecture is technically feasible for constrained applications because many of its constituent technologies already exist: secure execution environments, access-controlled APIs, cryptographic credentials and tokens, detailed audit mechanisms, privacy-preserving computation, policy enforcement and distributed data-processing infrastructure.

The more difficult questions are institutional and economic.

Will individuals entrust meaningful data to an independent custodian?

Will organisations accept authenticated answers rather than demanding possession of the underlying records?

Can authoritative data providers supply information to the Databank reliably?

Can regulators recognise the resulting outputs and institutional arrangements?

Can a Databank meet the costs of security, compliance, auditing, identity verification, customer support and dispute resolution without developing commercial incentives to exploit the information it holds?

A pilot is therefore needed not simply to demonstrate that the software can be built, but to test whether the surrounding **trust, governance, regulatory and business model** can function.

— 3. What the Scottish pilot would test

The pilot would seek to establish four forms of viability simultaneously.

Technical viability Can a Databank reliably derive and release useful information while preventing unnecessary disclosure of the underlying data?

For constrained use cases, the working hypothesis is that technical viability is high.

Institutional viability Can an independent custodian maintain a credible structural separation between custody and exploitation of personal data, supported by appropriate governance, auditability, security and legal duties?

Market viability Will individuals use such an intermediary, and will organisations accept its constrained outputs rather than collecting the corresponding raw information themselves?

Commercial viability is plausible but currently unproven.

Regulatory viability Can Databank outputs, controls, accountability mechanisms and institutional responsibilities be made sufficiently robust that regulators and relying organisations can treat them as trustworthy?

The model may be particularly suitable for testing in sectors that are already heavily regulated and in which the cost and liability associated with retaining personal data are significant.

— 4. A vertical-first approach

The Scottish pilot should deliberately avoid attempting universal personal-data custody.

The weakest assumption of a universal Databanking model is that a new institution could rapidly become trusted with every category of an individual's personal information. That is unnecessary for demonstrating the underlying architecture.

A more realistic path is to begin with one specialised sector in which:

- sensitive personal information is repeatedly requested;
- organisations need only a limited fact or decision derived from that information;
- possession of the underlying record imposes meaningful security, compliance or liability costs;
- authoritative sources of the relevant data already exist;
- the result has sufficient economic value to justify a transaction fee or other sustainable revenue model; and
- the benefits of minimising disclosure are intelligible to both the individual and the relying organisation.

Candidate domains include age and identity verification, financial eligibility and affordability, professional credentials, health and insurance information, research access to sensitive records, address and eligibility services, and controlled personal-data access for AI systems.

Age verification for alcohol sales offers a timely candidate. Since September 2026, licensed premises in England and Wales may accept certified digital proof of age, under which the customer shares only that they are 18 or over. The change does not extend to Scotland, which has its own licensing law, and the Scottish Government has announced a public services app that is to include digital proof of age. How a digital answer may stand in for a document at a Scottish counter is therefore a question Scotland will answer for itself.

The initial use case should be selected jointly with participating organisations and regulators. The technically simplest demonstration may not necessarily be the most useful pilot. Selection

should balance implementation tractability against the extent to which the case genuinely tests the distinctive value of Databanking.

— 5. The pilot ecosystem

A realistic Databanking transaction involves more than the individual and the querying organisation.

The pilot should therefore model an ecosystem involving four principal roles:

The individual, who authorises the use of their information and whose interests the custodial framework is intended to protect.

The authoritative data provider, which supplies or verifies the underlying information from which answers are derived.

The Databank, which maintains custody, identity binding, policy enforcement, computation, auditing and controlled disclosure.

The relying organisation, which submits an authorised query and accepts the resulting attestation or output instead of requiring possession of the underlying record.

Regulators and independent auditors would provide an additional governance layer.

The pilot would establish explicit responsibilities and liability boundaries between these participants rather than treating the Databank as an isolated technical platform.

— 6. Solving the initial adoption problem

Databanking has an inherent two-sided or multi-sided network problem.

Individuals may have little reason to join a Databank until organisations accept Databank queries. Organisations may have little reason to integrate with Databanks until a meaningful number of customers use them. Data providers may likewise have little incentive to build interfaces before demand exists.

A public launch followed by an attempt to grow each side independently would therefore be a weak test of the model.

The Scottish pilot should instead establish a **closed initial ecosystem**.

One or more relying organisations would agree in advance to accept a defined Databank output. Appropriate data providers would agree to supply or verify the necessary information. A bounded cohort of individuals would then be recruited for whom the supported transaction has an immediate practical use.

The pilot would consequently create the minimum network necessary to test the underlying proposition without requiring mass adoption beforehand.

If successful, subsequent phases could determine whether adoption can expand beyond this deliberately constructed initial network.

— 7. Technical and governance research programme

The demonstrator should sit alongside an interdisciplinary research programme rather than being treated simply as a software-development exercise.

Several questions require substantive research and engineering work:

1. **Query leakage.** Individually innocuous answers can collectively disclose sensitive information. Repeated threshold queries, for example, may progressively reconstruct an underlying value. The architecture therefore requires controls over query histories and cumulative information release, not simply per-query permission.
2. **Computation and algorithm scrutiny.** If external organisations can submit computations to a Databank, mechanisms are needed to ensure that these computations are safe, authorised and do not covertly extract information. Relevant approaches may include constrained query languages, static analysis, sandboxing, information-flow controls, output restrictions, privacy budgets, formal policy models and human review.
3. **Data correctness and provenance.** A perfectly secure Databank can still return an incorrect result if the underlying information is inaccurate or outdated. The system requires provenance, correction procedures and explicit responsibility for data quality.
4. **Identity binding.** The Databank must establish that records correspond to the correct individual and that authorisations originate from the appropriate person or authorised representative.
5. **Liability and dispute resolution.** Responsibility must be defined when errors or harmful outcomes result from the Databank, source data, submitted computation or relying organisation.
6. **Coercive authorisation.** Formal user consent does not necessarily imply meaningful choice. Employers, lenders, insurers, landlords or public authorities could potentially make excessive queries a de facto condition of access. The governance model must therefore regulate permissible requests as well as user authorisation.
7. **Security concentration and resilience.** Reducing uncontrolled duplication of personal information also concentrates valuable data. The pilot must investigate architectural compartmentalisation, resilience, breach containment and recovery.
8. **Interoperability and portability.** The long-term model should permit multiple competing Databanks rather than creating a single universal custodian. Data, policies and attestations must therefore be portable and interoperable across regulated providers.
9. **Economic sustainability.** The pilot should quantify the cost of identity verification, ingestion, storage, computation, security, compliance, audit, support and dispute resolution, and compare those costs with plausible revenues from custody and authorised access.

These questions naturally span computer science, cybersecurity, privacy-enhancing technologies, information law, regulation, economics, business models and public policy.

— 8. Success criteria

The decisive test of the pilot would be whether it demonstrates three propositions simultaneously.

First, **individuals are willing to entrust the relevant information to the custodian** under the proposed governance and accountability model.

Second, **relying organisations are willing to accept minimal authenticated outputs instead of obtaining the underlying personal records**.

Third, **the Databank can provide this service on economically sustainable terms while maintaining strict independence from commercial exploitation of the information under its custody**.

Secondary measures would include accuracy, transaction latency, security incidents, user understanding, dispute rates, data correction rates, organisational integration costs, compliance costs and the effectiveness of controls against excessive or cumulative disclosure.

Failure on any of these dimensions would itself provide useful evidence about where the Databanking model requires modification.

— 9. Why Scotland?

Scotland offers a potentially attractive environment for such a pilot because the problem is inherently interdisciplinary and requires collaboration across technology, regulation, academia, public policy and industry.

A Scottish pilot could bring together expertise in distributed systems, privacy-preserving computation, cybersecurity, data science, information law, financial technology, public-sector data governance and regulated services within a comparatively compact institutional ecosystem.

Scotland is also of a scale at which a bounded infrastructure experiment could be meaningful without requiring national-scale deployment from the outset.

The objective would not initially be to persuade Scotland to adopt Databanking universally. It would be to determine whether Scotland could serve as an **incubator for a new regulated data-infrastructure model**, establishing the technical architecture, governance mechanisms, regulatory principles, economic evidence and interoperability standards required to determine whether broader deployment is justified.

If successful, this could create expertise and intellectual property around a potentially important new infrastructure sector while providing an evidence base for future policy.

— 10. Proposed programme structure

A practical programme could proceed through several stages.

Phase 1 — Scoping and consortium formation Identify the strongest initial sector and transaction; recruit relevant academic, industry, public-sector and regulatory expertise; specify the

roles of the Databank, data providers and relying organisations; and establish the legal and economic assumptions to be tested.

Phase 2 — Architecture and governance design Develop the technical architecture, policy model, identity and authorisation mechanisms, query controls, audit framework, security model, liability structure, portability requirements and institutional governance.

This phase would also produce a more detailed Technical Architecture Paper.

Phase 3 — Prototype Implement a constrained prototype supporting a small number of standardised queries or attestations using synthetic and controlled test data.

The purpose would be to validate architecture and security assumptions before involving operational personal data.

Phase 4 — Regulated demonstrator Deploy the model with a bounded participant population, agreed data providers and one or more relying organisations.

The demonstrator would collect technical, behavioural, economic and governance evidence.

Phase 5 — Evaluation and expansion decision Evaluate whether the model has demonstrated sufficient technical, institutional, regulatory and commercial viability to justify expansion into additional queries, sectors or Databank providers.

A successful pilot should lead not automatically to scale-up, but to an evidence-based decision about which aspects of the model merit further development.

— 11. Longer-term direction

The intended destination is not a single institution holding all personal data.

A mature Databanking ecosystem would more plausibly contain **multiple regulated and interoperable Databanks**, potentially specialising in different sectors or services while operating under common standards for custody, identity, authorisation, policy enforcement, auditing and portability.

Individuals should be able to move between custodians without becoming locked into a particular provider.

Organisations should be able to request standardised outputs without needing bespoke integration with every Databank.

Regulatory obligations should apply consistently across providers.

The Scottish pilot could therefore contribute not only a working demonstrator but the beginnings of an **open institutional and technical framework for regulated personal-data intermediation**.

— 12. Proposed next step

At this stage, the proposal is to establish a small interdisciplinary group to examine the model, identify the strongest Scottish pilot use case and determine the appropriate institutional and funding structure.

The immediate requirement is therefore not a commitment to deploy Databanking at scale.

It is to bring together a small number of organisations and individuals with relevant expertise in technology, regulation, law, public policy and industry to answer a more bounded question:

Can Scotland demonstrate, in one carefully selected domain, that valuable decisions can be made using personal data without requiring the organisations making those decisions to possess the underlying personal record?

If the answer can be demonstrated convincingly, the resulting technical architecture, regulatory framework and economic evidence would provide a basis for deciding whether Databanking should develop into a broader infrastructure industry.