Welcome to the Digital Trace

A short introduction to what we are building and why.

Luca Franceschi and Jose Carrasco · digital-trace.org · *October 2026*

Our information is valuable: some of the largest companies in the world are built on it. Take, for instance, our countless web searches. They gave us useful results, and, likely, that was it for us. Instead, the parties receiving such information put it to many other uses, from serving targeted advertising to improving AI models, generating much value along the way. Yet they didn't "take the data from us", nor did we have to "give it up in exchange". Data is a *nonrival good*, one that many can use at once without anyone having less of it. So *we could hold it too.* Only, we don't.

Sure, some of our information stays with us: years of emails, folders of documents and photos, data in the apps we use. But it is scattered, of little use beyond its immediate context, hard to turn into evidence. Consider a product review we read. The platform may attach a "verified purchase" badge to it. The badge gives us some confidence in the review, but leaves open other important questions. Did they actually use the product? Did they write the review themselves? When we are the ones writing, the questions fall on us. In principle, *we* could settle them with evidence, but we lack the means to do so. Nobody can, so nobody expects it. Without such evidence, we are left trusting the platforms that serve us information rather than the people who produce it.

This is how, over the years, we have become *permanent guests* in our digital spaces. We hold little we can put to further use. We depend on our hosts for much, even for trust. The more we depend on them, the less we can do on our own. Yet the digital sphere is not a place we can simply leave. What we can do is stop being mere guests.

From guests to citizens

Three facts mark the way. Data we generate *comes from us* and is *ours to capture.* Once we *hold* it, we are free to *act* on it however we choose. Because all our activity *runs through us*, continuously, we are the best-placed party to *supply evidence* about it. All this may sound obvious, yet the means to do so have been missing, for evidence above all. We are building these means, all in one place:

> *The Digital Trace is open-source software you can run on each of your devices. It* ***captures*** *your digital activity privately, as it happens, and* ***keeps*** *it under your control. With it, you can* ***build*** *on that activity,* ***share*** *it on your terms,* ***produce verifiable claims*** *about it, and* ***check*** *those others make about theirs. It is free to use, integrate and extend.*

The Digital Trace doesn't stop parties you interact with from getting the information you send them. We don't think that's realistic, or even right. But by running the Digital Trace, you keep the originals, organized as *records* in one place, ready to serve many purposes at once: you can search your data whenever you need, feed it to models you own, pass it to friends and family for theirs, and license it to a dozen parties, each on terms you set.

Others don't have to take your word that what you hold is genuine: with the Digital Trace you can provide evidence that it is. Such evidence rests on cryptography and *origin attestation algorithms*: public programs that check whether the activity behind a record fits both human constraints (nobody types on three keyboards at once) and your own established patterns. You run them privately, as *verifiable computations*: each run yields a *zero-knowledge proof* that the program ran correctly, revealing no private content. Anyone can verify the proof independently without having to trust you (or, in general, whoever runs the computation).

In this way, the Digital Trace reconciles two needs that pull against each other: privacy and credibility. We are the party most interested in our privacy, so evidence about our claims must reveal nothing beyond what is strictly necessary. We are also the party most tempted to bend that evidence in our favour, so it must be costly for anyone, ourselves included, to fake, and checkable by

anyone, independently. What we can hold, then, goes beyond the copies other parties may have: a complete account, across time and activities, carrying evidence that it is genuine. Others see only slices (e.g. what we did while using their services), which carry no such evidence.

The Digital Trace is an *opportunity to create value*, and a *way to credit* those it came from. A party that credits you for using your records receives proof that it did so. By combining such proofs, it can show others that a model, a study, or a product it offers rests on information obtained on agreed terms. A party that uses your records without crediting you has nothing to show. Crediting people then becomes a rational choice, not a matter of goodwill alone.

Rooted in the real economy, not finance, the Digital Trace is neither a blockchain nor a cryptocurrency. There's no shared ledger, no global consensus, no tokens, no mining. You hold your records privately, linked into a privacy-preserving, tamper-evident chain: your *trace*. The trace is cross-checked by a few other participants, your *witnesses*, who verify its integrity and anchor it in time, but learn none of its content. The Digital Trace follows the principle of "*one person, one trace*": you have a single pseudonymous identity that stays with you for good. Your trace is valuable because it is the only one you have. Scarcity comes naturally, from people and what they produce.

The Digital Trace integrates smoothly into daily life and asks nobody to change their behaviour: you keep using platforms and services as before, organizations keep operating as they do. Nor does it trade privacy away for usefulness. Capture and storage are fully tunable, and you can delete record content at any time, except for a minimal cryptographic fingerprint. Disclosure is fine-grained and under your control: access is granted one request at a time, on conditions you state, and each grant leaves a receipt for both sides. Often you need not share records at all. When a party needs a result, such as a statistic or a score, you can share just that. You run the program you agreed on as a verifiable computation and send only the output and its zero-knowledge proof.

A few things you could do with the Digital Trace

- **Search and personal models.** At its simplest, the Digital Trace lets you search all your digital activity from one place, and answer questions like "where did I first come across this idea?". Beyond search, your records and those shared with or licensed to you can feed a recommender, an assistant, a personal digital memory, even an advertising model — not necessarily built from scratch, but e.g. fine-tuned or contextualized from public models. You can use these models yourself, or let others, platforms included, query them on your terms. From them, you may get better suggestions, personalization that meets your goals, financial returns, a voice of your own. Organizations get more accurate, up-to-date information about you, without holding your data.
- **Proofs of effort.** You can attach a checkable statement to anything you make (e.g., this piece took N hours and M revisions), backed by the origin attestation of the activity behind it, without showing the work itself. A freelancer can show a piece took days to draft; a student, that they wrote their homework; a reviewer, that they actually bought *and* used the product. In turn, you can check the proofs others attach, and weigh content that lacks one accordingly.
- **Data exchanges.** You can license your records (journeys, chatbot preferences, engagement with digital content), or results computed from them, through exchanges that match suppliers and consumers. You get something back for what you already produce (money, services, research you care about), and can change the terms, or stop, whenever you like. Data consumers get a large pool of natural, verifiable data, more representative than what tracking and surveys can collect.

Why now?

Needed now. Living as guests in the digital sphere is not a new condition, but in just a few years AI has made the issue more pressing. As automated agents, often malicious, populate the internet, we increasingly face content and interactions that are synthetic and hard to tell from genuine ones. On top of this, AI trained on what we produce is taking on a growing share of what we do. We have little say in how this happens, and risk being cut out of the value it creates. Origin, attribution, agency are no longer optional: they are how we keep our place in the digital sphere.

Valuable now. AI also brings opportunity. For the past decade, value lay in data at scale. Now the typical is cheap to access, while the idiosyncratic (particular tastes, behaviour, expertise) remains scarce, and hence valuable. So do human sources, as the web fills with generated content. AI developers need them: models trained without enough human data degrade, or "collapse". AI tools also make it more practical for anyone to process their own data at scale, increasing the usefulness of raw material to whoever holds it. Regulation, in the EU above all, is closing in on current business intelligence practices, leaving room for data people supply themselves, on their own terms.

Feasible now. The building blocks are ready, or nearly so. Zero-knowledge virtual machines, which implement verifiable computation, already run on consumer hardware for simple programs, with room left to make them cheaper and more expressive. Witnessing, building on the transparency logs that secure web certificates, removes the need for global state and consensus, keeping running costs low and letting the system scale with its participants and their devices.

What the Digital Trace changes

With the Digital Trace, we turn our digital activity and interactions into a resource we hold, can learn from, and can choose to put to work. A journey taken today may back a review tomorrow, improve a personal model next month, and inform urban planning next year. A music recommender of our own can draw not only on what we listened to or what is popular at the moment, but also on places, people, and activities, past and present. We can plug it into the streaming services we use, or lend it to a friend for the night. We become *digitally richer* not because we consume more, but because we gain something to show for our activity, and the raw material to build on.

As the Digital Trace spreads, incentives shift towards honesty. People who produce genuine work can present evidence of its origin. Parties that obtain genuine records on agreed terms get receipts they can show. Since this evidence is cheap for honest parties to produce, expensive to fake, and easy for anyone to verify independently, people come to expect it. Its absence speaks for itself: rather than chasing offenders, we can simply *prefer* those with the right evidence. As a result, our societies also become more resilient to information attacks since attackers cannot multiply identities at will, and the few identities they may hold cannot back content beyond human pace.

With trust in place, what we hold becomes far more productive, whether records, results computed from them, or access to our personal models. In our economies, information is a factor of production. As individuals freely balance what they disclose, directly or through exchanges, against what they get for it, more data enters production. The information itself also gets better. It arises from routine activity, in natural contexts, and is more representative than anything tracking or surveys can gather. It is information we have every reason to curate: correcting a model of our own, which we hold and benefit from, makes far more sense than answering a platform's feedback request.

Organizations of every kind benefit. New companies can find what they need in open exchanges from day one, and save on storage and compliance costs. Established organizations gain access, with consent, to first-hand information beyond the reach of their services: they can obtain richer feedback on what we do with their products and recommendations, and natural signals that help them improve. They can hold less personal data and ask for less of it upfront, receiving only the results they need, each with a zero-knowledge proof: a practical approach to data minimization. Researchers and public bodies can reach information that is today too fragmented, sensitive, or costly to obtain. AI developers obtain, with permission, continuing streams of genuine human activity. New economic and research activity becomes possible thanks to more complete and accurate information, and to a kind of trust, grounded in evidence, that no single party can provide today.

The larger promise of the Digital Trace is not merely better privacy, nor merely a new market for data, but *firmer ground in our digital lives*. A world in which we can act on what we produce, organizations have better means to create value and to credit the people behind it, and contributions can be traced, so that trust and reward follow them. A world in which every party stands to gain as we become citizens, not guests, of our digital spaces.

Where we are

This document is a short introduction to the Digital Trace. A comprehensive working paper, available on request, sets out extended arguments, further description of possible applications, and a technical specification. We have built a prototype covering capture, storage, and encryption, with a desktop client for macOS, Windows, and Linux (Ubuntu/Debian). This is also available on request. We are now formalizing the technical specification in full, and building, in Rust, a first complete version of the Digital Trace with every core part in place, some deliberately simplified. We are looking for collaborators and funders to bring this project forward.

Reach us at welcome@digital-trace.org

References

Our account of automated agents draws on Imperva (2026), our economic arguments (the nonrivalry of data and its implications) on Jones and Tonetti (2020), witnessing on transparency logs and cosigning (Laurie, 2014; Syta et al., 2016), verifiable computation on zero-knowledge virtual machines (Bruestle et al., 2023; Succinct Labs, 2024), our account of regulation on European Parliament and Council of the European Union (2022, 2023), and our account of model collapse on Shumailov et al. (2024); Gerstgrasser et al. (2024); Dohmatob et al. (2025).

Jeremy Bruestle, Paul Gafni, and the RISC Zero Team. RISC Zero zkVM: Scalable, transparent arguments of RISC-V integrity. Technical report (draft), 2023. URL `https://dev.risczero.com/proof-system-in-detail.pdf`. Accessed: 2026-08-25.

Elvis Dohmatob, Yunzhen Feng, Arjun Subramonian, and Julia Kempe. Strong model collapse. In *International Conference on Learning Representations (ICLR)*, 2025.

European Parliament and Council of the European Union. Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector (Digital Markets Act). `https://eur-lex.europa.eu/eli/reg/2022/1925/oj`, 2022. OJ L 265, 12.10.2022, p. 1–66; see Art. 5(2).

European Parliament and Council of the European Union. Regulation (EU) 2023/2854 on harmonised rules on fair access to and use of data (Data Act). `https://eur-lex.europa.eu/eli/reg/2023/2854/oj`, 2023. OJ L, 2023/2854, 22.12.2023; see Arts. 4–5.

Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang, David L. Donoho, and Sanmi Koyejo. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. *arXiv preprint arXiv:2404.01413*, 2024.

Imperva. Bad Bot Report 2026: The internet is no longer human and it's changing how business works. `https://www.imperva.com/blog/bad-bot-report-2026-bots-agentic-age/`, April 2026. Accessed: 2026-07-27.

Charles I Jones and Christopher Tonetti. Nonrivalry and the economics of data. *American economic review*, 110(9):2819–2858, 2020.

Ben Laurie. Certificate transparency. *Communications of the ACM*, 57(10):40–46, 2014.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. Ai models collapse when trained on recursively generated data. *Nature*, 631(8022):755–759, 2024.

Succinct Labs. Introducing SP1: A performant, 100% open-source, contributor-friendly zkVM. Blog post, 2024. URL `https://blog.succinct.xyz/introducing-sp1/`. Documentation at `https://docs.succinct.xyz/docs/sp1/introduction`. Accessed: 2026-08-25.

Ewa Syta, Iulia Tamas, Dylan Visher, David Isaac Wolinsky, Philipp Jovanovic, Linus Gasser, Nicolas Gailly, Ismail Khoffi, and Bryan Ford. Keeping authorities "honest or bust" with decentralized witness cosigning. In *IEEE Symposium on Security and Privacy (S&P)*, pages 526–545, 2016.