

How We Work With Evidence

A methodology for turning questions into verifiable, usable evidence.

QUESTION	Define what must be established
EVIDENCE	Map and collect relevant sources
STRUCTURE	Turn records into comparable data
ANALYSE	Find patterns, exceptions and relationships
VERIFY	Test findings and competing explanations
LIMITS	State what the evidence cannot establish

**Research at Dataful begins with
a question, not a dataset**



OUR APPROACH

Evidence is built, tested and made traceable.

We start by defining what we are trying to understand, what evidence would be needed to answer the question, and what would constitute a sufficiently reliable basis for a finding. Only then do we begin looking for data, documents and other forms of evidence.

This matters because information is rarely evidence on its own. A database may contain thousands of records but still leave the central question unanswered. A government document may establish that something was officially recorded but not that it happened exactly as described. A company disclosure may provide important facts while reflecting only what the company is required or willing to disclose.

Our research therefore focuses on building chains of evidence rather than accumulating sources.

**The aim is not to make research appear certain.
It is to make the basis of a conclusion visible.**

**QUESTION**

Define what must be established

We begin by defining the question

Every investigation or research project starts by turning a broad subject into questions that can be tested against evidence. Instead of beginning with What information can we find?, we ask: What are we trying to establish? What would we need to know to establish it? What evidence should exist if a particular explanation is correct? This determines what data we collect, which documents we search for, what variables we construct and where we look for corroboration.

The question may change as new evidence emerges. When that happens, we document the change rather than forcing the evidence into the assumptions with which the research began.

**EVIDENCE**

Map and collect relevant sources

We identify the evidence that can answer it

Once the question is defined, we map the evidence environment around it. Depending on the research, this may include government records, administrative data, corporate disclosures, regulatory filings, procurement records, court documents, international databases, technical registries, archived webpages, satellite or geospatial information, social-media data, surveys, interviews and previously published research. Different sources perform different functions. Some establish identity.

Others establish dates, transactions, relationships, quantities, ownership, behaviour or institutional responsibility. Some provide context but cannot independently substantiate a finding. We therefore do not treat every source as equivalent. We ask what each source is capable of proving and, equally importantly, what it cannot prove.

We establish stable points of reference

Names, organisations, projects and events can appear differently across datasets and documents. They can be renamed, translated, abbreviated, misspelled or recorded inconsistently. Where possible, we therefore identify a stable reference around which evidence can be organised. This might be a company registration number, procurement identifier, vessel IMO number, geographic coordinate, contract number, account identifier, date, case number or another persistent attribute.

We then use that reference to connect records that would otherwise appear unrelated. This process of entity resolution is particularly important when research crosses multiple databases, institutions, countries or periods of time.

We build the evidence across sources

Important findings rarely depend on a single document. We reconstruct the subject by bringing together evidence produced for different purposes and by different institutions. A regulator may establish one part of a history. A corporate filing may establish another. Administrative data may reveal a pattern that neither describes directly. An international database may allow an identity or transaction to be checked independently.

The objective is not simply to find several sources saying the same thing. It is to determine whether independent pieces of evidence fit together. Where they do not, the disagreement itself becomes something to investigate.



STRUCTURE

Turn records into comparable data

We structure information into data

Much of the evidence needed for research does not arrive as a clean dataset. It may exist across hundreds of PDFs, webpages, spreadsheets, reports, filings or individual records. We extract the relevant information and convert it into structured data so that records can be compared systematically. During this process, we define variables, standardise names and dates, identify duplicates, document exclusions and preserve links to the underlying evidence wherever possible.

When classifications or analytical categories are created by Dataful rather than supplied directly by the source, we distinguish those derived variables from original source data. This allows us to move from individual records to patterns without losing the connection between an analytical finding and the evidence from which it was derived.



ANALYSE

Find patterns, exceptions and relationships

We analyse patterns, exceptions and relationships

Once the evidence has been structured, analysis begins. The specific techniques depend on the question. We may examine changes over time, distributions, concentrations, relationships between entities, geographic patterns, networks, financial flows, differences between groups, anomalies or recurring behaviours. Quantitative analysis helps us identify patterns. It does not automatically explain them. We therefore return repeatedly from the aggregate finding to the underlying records.

An unusual number, concentration or relationship is examined in its original context before we decide what it means. Where statistical, computational or automated methods are used, their outputs are treated as analytical evidence requiring interpretation rather than as conclusions in themselves.

We distinguish evidence from inference

Not every conclusion is stated directly in a document or dataset. Research often requires inference: several independently established facts may together support a conclusion that no individual source explicitly states. When we make such an inference, we try to make the reasoning visible. We distinguish between what a source directly establishes, what our analysis shows, and what we infer from the combination of those findings.

The stronger the claim, the stronger we expect the supporting chain of evidence to be.



VERIFY

Test findings and competing explanations

We test alternative explanations

Finding evidence that supports an explanation is only part of the research process. We also look for evidence that could contradict it. We ask whether the observed pattern might result from incomplete data, changes in reporting practices, duplicated records, differences in definitions, selection effects or another plausible explanation. Where possible, important findings are tested using an independent source, method or dataset.

This adversarial approach to our own findings helps distinguish what the evidence actually demonstrates from what merely appears plausible.

We verify before we conclude

Verification is not a final step added after analysis. It runs throughout the research. Names are checked against identifiers. Dates are compared across sources. Numerical totals are recalculated. Important records are traced back to their original source. Findings produced from structured datasets are checked against individual records.

For consequential findings, we seek corroboration from evidence that is independent of the source on which the original finding depends whenever such evidence is available. If two credible sources conflict, we do not silently choose one. We investigate the reason for the discrepancy and disclose unresolved contradictions when they materially affect the finding.

**LIMITS**

State what the evidence cannot establish

We document what the evidence cannot tell us

A methodology should explain not only how a conclusion was reached but also where the evidence stops. Public records can be incomplete. Databases can omit cases. Documents can remain confidential. Historical webpages can disappear. Corporate disclosures may contain only information required by regulation. Social-media datasets may represent only observable activity rather than the full information environment. We therefore identify the limits of the available evidence.

If we cannot establish something, we say so. If the evidence supports an association but not causation, we do not describe it as causal. If available records cover only part of a period or population, we do not generalise them to the whole. If an important document is unavailable, we identify that absence rather than filling the gap with assumption. For us, what remains unknown is part of the finding.

We preserve the route back to the evidence

Our research is designed so that findings can be traced backwards. A reader should be able to move from a conclusion to the analysis behind it, from the analysis to the underlying records, and where possible from those records to the original public source. For major investigations, we may publish datasets, sourcebooks, methodological notes, evidence tables or explanations of how the research was conducted. This does not mean that every source can always be made public.

Ethical, privacy, security, contractual or legal considerations may restrict publication of some material. Where that happens, we explain the restriction when doing so is possible and appropriate.



EVIDENCE IS THE METHOD

What we found - and how we know.

The purpose of this process is not to make research appear certain. It is to make the basis of our conclusions visible.

We find and structure data. We connect records across sources. We analyse what they show. We test findings against alternative explanations. We verify important claims independently. And we identify the point beyond which the available evidence does not allow us to go.

That is how Dataful turns questions into usable evidence.

And by making that process visible, we make it possible for others to examine not only what we found, but how we know.