Government Impact Index (GII)

A Causal, Multi-Institution Framework for Tracing the Effects of United States Government Decisions

Research Paper — Methodology Foundation

VERSION	1.1.1 • July 19, 2026
AUTHOR	Kurt Sather

Conceptual, methodological, and research foundation

The v1.1 revision line preserves the v1.0.0 foundation and adds the initial PEPI source-adapter standard and formal contribution credits. It does not publish or imply production GII scores for any administration, party, branch, court, agency, law, regulation, or policy.

Contents

The document is organized for both decision-level reading and technical implementation.

- Abstract
- Research Summary
- **1. Introduction**
- **2. Institutional and Measurement Scope**
- **3. Research and Practice Foundations**
- **4. Conceptual Model: From Decision to Impact**
- **5. Unit of Analysis and Decision Impact Record**
- **6. Domain Measurement Framework**
- **7. Causal Identification and Evidence Grading**
- **8. Index Construction**
- **9. Data Provenance, Reproducibility, and Governance**
- **10. Public-Data-First Research Architecture**
- **11. Validation Program and Research Propositions**
- **12. Limitations and Failure Modes**
- **13. Development Roadmap**
- **14. Conclusion**
- References

Version 1.1.1 • PEPI integration baseline

Abstract

Government decisions affect households, markets, institutions, and civil society through legal, administrative, fiscal, monetary, and behavioral channels. Existing public systems document what government has enacted, decided, spent, regulated, or adjudicated, while specialized evaluations estimate selected fiscal, economic, health, environmental, or program outcomes. These resources are valuable but fragmented. They generally do not provide a unified, continuously updated method for linking a decision made by a specific United States government institution to its implementation, exposure, downstream outcomes, affected stakeholders, and evidentiary confidence.

This paper proposes the Government Impact Index (GII), an event-based, causal, multi-institution measurement system. GII is designed to trace the effects of federal government decisions across six primary impact domains: people and households; the United States economy; public entities; nonprofit and civil-society organizations; international trade; and private enterprise. The framework recognizes Congress, presidential actions, executive departments and agencies, independent regulators, the federal judiciary, the Supreme Court as a separately reportable judicial subindex, and the Federal Reserve as distinct decision origins. It represents statutes, rules, orders, court decisions, appropriations, enforcement actions, monetary decisions, and their implementing events as linked policy lineages rather than isolated documents.

The central methodological choice is to measure impact without forcing unlike consequences into one partisan or moral ranking. The primary GII score measures attributable impact magnitude on a 0–100 scale. A separate directional balance reports the measured mix of benefits and burdens from −100 to +100 using preregistered outcome-orientation rules. Attribution confidence, coverage, implementation status, timing, and distributional concentration are published alongside every score. Forecast impacts and observed impacts are never merged. Rights, legal access, institutional capacity, and other consequences that resist monetization remain visible and are not treated as zero.

GII combines official decision records, administrative and statistical data, peer-reviewed causal methods, transparent normalization, uncertainty analysis, and versioned data provenance. The proposed result is not a score of whether government is “good” or “bad.” It is an auditable infrastructure for answering a narrower and more useful question: what changed, for whom, by how much, over what period, and how confidently can the change be traced to a government decision?

Keywords: government impact; public policy evaluation; causal inference; composite indicators; United States government; judicial impact; regulatory analysis; distributional analysis; policy provenance; evidence-based policymaking

Research Summary

The Government Impact Index is intended to fill five related measurement gaps.

1. **Institutional fragmentation.** Legislative, executive, regulatory, judicial, and monetary decisions are recorded in different systems and evaluated under different conventions.
2. **Decision-to-outcome fragmentation.** Official records usually establish that an action occurred, while outcome datasets measure conditions without a durable causal link back to the action.
3. **Forecast-versus-observation confusion.** Prospective cost estimates and regulatory analyses are often discussed as though they were realized outcomes.
4. **Aggregation without distribution.** National totals can conceal concentrated gains, losses, administrative burdens, or rights effects.
5. **False certainty.** A single score can hide weak identification, incomplete coverage, revised source data, and contestable weights.

GII addresses these gaps through a versioned Decision Impact Record, a policy-lineage graph, an explicit causal design, and a multidimensional scorecard. Its production architecture is public-data-first and API-first. Government publications establish legal and administrative facts; official statistics and audited datasets establish outcomes; peer-reviewed methods estimate counterfactual effects; and every published value retains its source, data vintage, model version, assumptions, and review state.

Version 1.1.0 change note

This minor revision is backward-compatible with v1.0.0. It does not change the GII construct, core formulas, institutional taxonomy, six impact domains, causal-evidence standard, or separation of magnitude from direction. It adds a defined adapter for the Presidential Economic Pressure Index (PEPI) at pepi.apiautomations.com, establishes rules for using PEPI records as upstream evidence without importing PEPI's composite score into GII, and records the concept and ideation credits below. The v1.0.0 publications remain preserved as immutable prior versions.

Version 1.1.1 change note

Version 1.1.1 is a patch revision under the GII semantic-versioning policy: clarifications and nonbreaking corrections only. It corrects the distributional-concentration share convention so both suite documents state the fractional-share (total variation) form; defines the oriented-indicator set for directional balance; designates the operational A–F evidence ladder as the suite's normative scale; harmonizes scorecard symbols, figure numbering, reference entries, and the Government Decision Event term across the suite; names the preregistration mechanism, inter-rater reliability statistics, licensing terms, and persistent-identifier policy; adds first-party source and contributor-relationship disclosures;

removes upstream identifier formats from the PEPI crosswalk in favor of neutral source referencing; and restates authorship in a research context. No construct, formula weight, threshold, taxonomy, or lifecycle definition changes.

Concept and Contributions

- **Author and methodology lead — Kurt Sather.** Research design, GII product architecture, causal framework, scoring methodology, and publication specification.
- **Index concept credit — Randall Hall, CEO, BookedByFriday.** Credited for the index concept that prompted development of the Government Impact Index.
- **Contributing ideation — Shivani D. Patel, CTO, API Automations.** Contributed early marketplace-oriented ideas that sharpened the framework's intended benefit across private, public, and consumer marketplaces.

These acknowledgments identify the nature of each contribution and do not imply co-authorship, review endorsement, or responsibility for methodological claims unless separately documented. Credited contributors may hold separate commercial or organizational relationships with the author; credits reflect intellectual contribution only.

1. Introduction

The United States government does not act through one institution or one instrument. Congress legislates, appropriates, taxes, and conducts oversight. Presidents issue orders, proclamations, memoranda, and directives. Executive departments and independent agencies promulgate rules, administer programs, award funds, enforce law, and sometimes adjudicate disputes. Federal courts interpret law, invalidate or sustain government action, and alter the practical rights and obligations of governments, organizations, and people. The Supreme Court sits within the judicial branch but can have distinct nationwide and precedential effects. The Federal Reserve makes monetary and regulatory decisions under congressionally assigned authority with consequences across employment, prices, credit, finance, trade, and business investment.

The constitutional branches are distinct but their functions are not empirically separable in every policy result. A statute may authorize a program, an agency may write implementing rules, states may administer benefits, a court may limit eligibility, Congress may revise funding, and later agency guidance may change delivery. Constitution Annotated describes the Constitution as delineating the roles of the legislative, executive, and judicial branches, while also recognizing that the practical exercise and

interpretation of authority involves interaction among them (Congressional Research Service, 2024). GII therefore treats institutional origin as a factual classification and policy impact as a lineage problem.

The federal evidence ecosystem already contains important foundations for this work. The Foundations for Evidence-Based Policymaking Act of 2018 established requirements for federal evidence-building, evaluation, data management, and open government data (Public Law 115-435, 2019). Office of Management and Budget guidance identifies rigor, relevance and utility, independence and objectivity, transparency, and ethics as central evaluation standards (OMB, 2020). The Government Accountability Office distinguishes performance measurement from evaluation and emphasizes that causal questions require evidence beyond routine output counts (GAO, 2021). CBO estimates the budgetary effects of most committee-approved bills and identifies intergovernmental and private-sector mandates, while regulatory analyses organize anticipated benefits and costs for selected rules (CBO, 2026a; OMB, 2003/2025).

These systems are not a general cross-government impact index. CBO estimates are primarily prospective and legislatively scoped. Regulatory impact analysis covers a defined subset of executive actions and is produced before or around implementation. GAO and agency evaluations are program-specific. Court databases describe cases, holdings, votes, dockets, or caseloads but rarely link decisions to downstream socioeconomic outcomes. Governance indices generally compare countries or institutional quality rather than estimate the effects of individual decisions. Policy-agenda datasets measure attention and activity rather than attributable real-world impact.

GII is proposed as a bridge between these layers. Its goal is to create a reproducible chain from the authoritative decision record to implementation, actual exposure, measured outcomes, and affected stakeholders. The project is grounded in causal inference rather than temporal coincidence, in transparent composite-indicator practice rather than opaque ranking, and in institutional provenance rather than narrative attribution.

1.1 Primary research question

How can the material, institutional, legal, social, and economic effects of United States government decisions be measured over time and traced to their decision origins without collapsing uncertainty, distribution, and qualitatively different outcomes into a misleading single judgment?

1.2 Supporting questions

- What constitutes a scoreable government decision event?
- How should related actions across branches and agencies be linked without double counting?

- Which outcomes are appropriately directional, and which should be magnitude-only?
- How should anticipated impacts be separated from observed impacts?
- What causal designs are credible for nationwide, staggered, threshold-based, judicial, regulatory, fiscal, and monetary interventions?
- How should affected people, public bodies, nonprofits, trade relationships, and companies be represented when the same decision benefits some groups and burdens others?
- What minimum evidence, coverage, and reproducibility standards should govern publication?

1.3 Proposed contribution

GII contributes a shared unit of analysis, a decision-origin taxonomy, a policy-lineage model, six stakeholder impact domains, a causal evidence ladder, a modular scorecard, a revision protocol, and a public-data-first source architecture. The intended product is simultaneously a research index, a decision registry, an evidence graph, and an API-ready public data system.

2. Institutional and Measurement Scope

2.1 Initial scope

Version 1.1.1 is designed for decisions of the United States federal government. The impact universe extends beyond federal institutions because federal actions are commonly implemented by or transmitted through state, local, tribal, and territorial governments; public schools, hospitals, universities, utilities, and transit systems; nonprofit organizations; households; firms; financial markets; and international trade relationships.

State and local decisions are not independent decision origins in the initial national index unless they are part of a federal implementation design or a comparison group used for causal identification. A later methodology version may add state and local subindices. This boundary prevents an unmanageable first release and avoids mixing constitutional levels of government before the federal lineage model is validated.

The people-and-households domain covers U.S. citizens as a specific stakeholder group while also measuring residents and households when federal outcomes, statistical denominators, or legal effects cannot validly be limited to citizenship. Citizenship status is retained as a distributional attribute where lawful, available, and analytically appropriate; it is not inferred at the individual level.

2.2 Decision-origin taxonomy

Code	Decision origin	Illustrative instruments	Treatment in GII
LEG-CONG	Congress	Public laws, appropriations, tax provisions, resolutions with legal effect, overrides	Legislative subindex and lineage origin
EXE-PRES	President and Executive Office	Executive orders, proclamations, presidential memoranda, directives, vetoes	Presidential-action subindex
EXE-AGY	Executive departments and agencies	Final rules, guidance with material operational effect, grants, procurement, enforcement, program administration	Executive-agency subindex
IND-REG	Independent commissions and regulators	Rules, orders, adjudications, enforcement, market-structure actions	Independent-regulatory subindex; institution tagged
JUD-SCOTUS	Supreme Court of the United States	Merits opinions, material orders, stays, injunction-related decisions	Standalone Supreme Court subindex nested within Judiciary
JUD-COA	Federal courts of appeals	Precedential opinions, nationwide or circuit-wide injunctions and stays	Appellate judicial subindex
JUD-DIST	Federal district and specialty courts	Material judgments, injunctions, vacatur, nationwide or high-reach relief	Lower-court subindex; geographic/legal reach explicit
FED-MON	Federal Reserve System	FOMC target decisions, balance-sheet policies, material supervisory or payment-system rules	Separate monetary/central-bank decision origin
INTG-IMPL	Intergovernmental implementer	State, tribal, territorial, and local implementation of federal policy	Implementation actor, not a fourth constitutional branch

The Supreme Court is not added twice. Its results are separately visible and then included once in the aggregated judicial score. The Federal Reserve is separately reportable because its operational independence, statutory mandate, decision cadence, and transmission channels make it analytically distinct, even though its authority ultimately derives from federal statute. Independent regulators are likewise distinguished from cabinet departments without implying that they exist outside the constitutional structure.

2.3 Included decision events

A decision enters the observed-impact pipeline when it has an authoritative record, a legally or operationally meaningful status, a plausible material pathway, identifiable exposure, and at least one measurable or systematically assessable outcome. Included event types may include:

- enacted statutes and provisions;
- appropriations and rescissions;
- tax changes;
- final rules, deregulations, and rescinded rules;
- executive orders and other presidential actions with operational effect;
- material agency guidance, waivers, enforcement shifts, procurement, grants, and administrative orders;
- Supreme Court and lower federal-court decisions that alter governing law or implementation;
- trade measures, sanctions, export controls, and tariff actions;
- Federal Reserve monetary-policy decisions and material regulatory actions; and
- implementation events necessary to activate a prior legal decision.

Speeches, campaign promises, press speculation, introduced bills, proposed rules, pending litigation, and nonbinding political statements may be retained in a forecast or watch registry but do not receive an observed GII score. A vote, announcement, or filing is not an impact merely because it occurred.

2.4 Six primary impact domains

Domain	Core question	Illustrative outcome families
People and households	How did lived conditions, resources, services, safety, rights, access, or administrative burden change?	Disposable income, prices, employment, health, safety, eligibility, service access, rights coverage, time burden
U.S. economy	How did aggregate and regional economic conditions change?	Output, employment, wages, inflation, productivity, investment, fiscal position, financial conditions
Public entities	How were public institutions and service providers affected?	Revenue, expenditure, mandates, staffing, capacity, service quality, legal exposure, capital needs
Nonprofits and civil society	How did mission-driven organizations and the people they serve experience the decision?	Grants, donations, demand, compliance, workforce, program capacity, mission access
International trade	How were cross-border flows, prices, investment, and supply chains affected?	Imports, exports, tariff incidence, trade balance, customs revenue, FDI, retaliation, supply-chain exposure

Domain	Core question	Illustrative outcome families
Private enterprise	How were firms and markets affected, including differences by size and industry?	Revenue, costs, compliance, labor, capital access, competition, innovation, entry/exit, investment

These domains are impact destinations, not policy topics. A health statute may affect all six domains; a court decision classified under civil rights may also change employer costs, state administrative systems, and household access. GII therefore stores both a policy-topic classification and one or more stakeholder impact domains. The Comparative Agendas Project's consistent 21-topic/220-subtopic scheme, Congress.gov policy areas and legislative subject terms, and Federal Register indexing terms offer useful starting taxonomies, but none substitutes for the impact-domain model (Comparative Agendas Project, 2026; Library of Congress, 2026).

3. Research and Practice Foundations

3.1 Evidence-based policymaking

The Evidence Act supplies a statutory rationale for systematic evaluation, learning agendas, data stewardship, and open government data. It does not require one cross-institution index, but it demonstrates that federal decision-making and accountability increasingly depend on structured evidence. GAO's synthesis identifies interrelated practices for planning results, assessing and building evidence, using evidence, and fostering learning and accountability (GAO, 2023). GII adapts these principles for an independent, public-facing measurement system.

Performance measures and impact evaluation must remain distinct. Outputs such as dollars obligated, inspections completed, claims processed, rules published, or cases decided can demonstrate implementation. They do not, by themselves, show that a policy caused an outcome. GAO defines evaluation as systematic data collection and analysis directed to questions of effectiveness and efficiency and highlights the need to match evaluation design to the question (GAO, 2021). GII therefore treats implementation outputs as mechanism evidence and outcomes as separate objects.

3.2 Causal inference

The causal problem can be stated using the potential-outcomes framework: for a unit, group, jurisdiction, firm, or time period, the effect of a decision is the difference between the outcome under the policy and the outcome that would have occurred without it. Only one state is observed, so the counterfactual must be designed or estimated (Rubin, 1974; Imbens & Wooldridge, 2009).

Government decisions are especially difficult treatments. They often respond to deteriorating conditions, apply nationally, begin at different times across jurisdictions, interact with other policies, produce anticipation effects, and change behavior before formal effective dates. Conventional before-and-after comparisons are therefore insufficient. GII uses a method-assignment framework that includes randomized or quasi-random implementation where available, difference-in-differences with modern staggered-treatment estimators, synthetic controls, regression discontinuity, instrumental variables, interrupted time series, event studies, panel and time-series models, microsimulation, structural models, and transparent contribution analysis.

Modern difference-in-differences research shows that standard two-way fixed-effects estimators can be biased when treatment timing and effects vary. GII requires explicit treatment-cohort and event-time analysis rather than automatic reliance on a single fixed-effects coefficient (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021). Synthetic controls are useful for a single treated jurisdiction or national policy when a credible weighted comparison can be built (Abadie, Diamond, & Hainmueller, 2010). Regression discontinuity can exploit statutory eligibility thresholds (Lee & Lemieux, 2010). Interrupted time-series designs can assess clearly timed population interventions when serial correlation, seasonality, concurrent events, and pretrends are addressed (Lopez Bernal, Cummins, & Gasparrini, 2017). Financial event studies can measure short-window market repricing, but market value is an early signal rather than a complete measure of social welfare (MacKinlay, 1997).

3.3 Regulatory, budgetary, and distributional analysis

OMB Circular A-4 describes regulatory analysis as a formal organization of evidence about expected consequences and provides conventions for benefits, costs, alternatives, uncertainty, and distributional effects. The 2023 revision was revoked in 2025 and the September 2003 circular was reinstated, illustrating why GII must store the applicable guidance version rather than assume analytic rules are politically or administratively permanent (OMB, 2003/2025; OMB, 2025).

CBO estimates provide disciplined baselines for legislative fiscal analysis and identify intergovernmental and private-sector mandates. They are advisory forecasts rather than observed causal estimates and must be stored as such (CBO, 2023, 2026a). CBO's distributional work demonstrates that policies can affect groups differently and that household results may be examined by income, wealth, age, education, race and ethnicity, and sex (CBO, 2022, 2026b). GII generalizes the principle of distributional disclosure to all six domains.

Benefit-cost analysis is useful for organizing favorable and unfavorable consequences, yet even leading proponents caution that uncertainty, fairness, process, and non-economic considerations matter (Arrow

et al., 1996). GII uses monetized values where defensible but does not force rights, dignity, institutional legitimacy, legal access, mortality, ecological effects, or public-service capacity into dollars merely to achieve aggregation.

3.4 Composite-indicator construction

Composite indicators are communication tools, but their apparent simplicity can conceal choices about theory, normalization, missingness, weighting, and aggregation. The OECD/JRC handbook recommends a clear theoretical framework, careful indicator selection, multivariate analysis, transparent normalization and weighting, sensitivity testing, and decomposition back to underlying indicators (OECD/European Union/EC-JRC, 2008). Saisana, Saltelli, and Tarantola (2005) show that uncertainty and sensitivity analysis are essential to assess whether rankings or policy messages are robust.

GII responds by making the composite subordinate to the evidence record. Every headline value must be decomposable to event, domain, outcome, group, source, estimate, and version. Alternative weighting and normalization specifications are published as robustness ranges. Missing data are not silently converted to average performance.

3.5 Legal measurement and judicial impact

Court opinions are authoritative legal events, not ready-made impact measures. Supreme Court slip opinions are replaced by edited United States Reports versions, and lower-court opinions vary in publication, precedential status, geographic reach, relief, and later treatment (Supreme Court of the United States, 2026; U.S. GPO, 2026). The Federal Judicial Center's Integrated Database, U.S. Courts statistics, GovInfo opinions, CourtListener, and the Supreme Court Database provide complementary metadata and research infrastructure (FJC, 2026; U.S. Courts, 2026; Free Law Project, 2026; Supreme Court Database, 2026).

GII does not use an ideological coding of a judicial decision as its benefit/burden direction. The legal holding, scope, precedential reach, implementation status, and affected legal interests are recorded as facts. Direction is assigned only to preregistered outcome measures—for example, measurable access, administrative cost, health, income, market structure, or legally defined coverage—and is left unrated where the welfare orientation is genuinely contestable. Legal epidemiology provides a useful model: law can be converted into structured data and evaluated as a causal factor, but credible measurement requires transparent legal coding and rigorous outcome analysis (CDC, 2024).

3.6 Gap addressed by GII

No reviewed framework jointly provides all of the following: cross-institution decision coverage; authoritative legal provenance; linked implementation events; causal attribution; six stakeholder domains; forecast/observation separation; distribution; confidence; revision history; and a decomposable national time series. This is a qualified design-gap finding, not a claim that no related index or evaluation exists.

4. Conceptual Model: From Decision to Impact

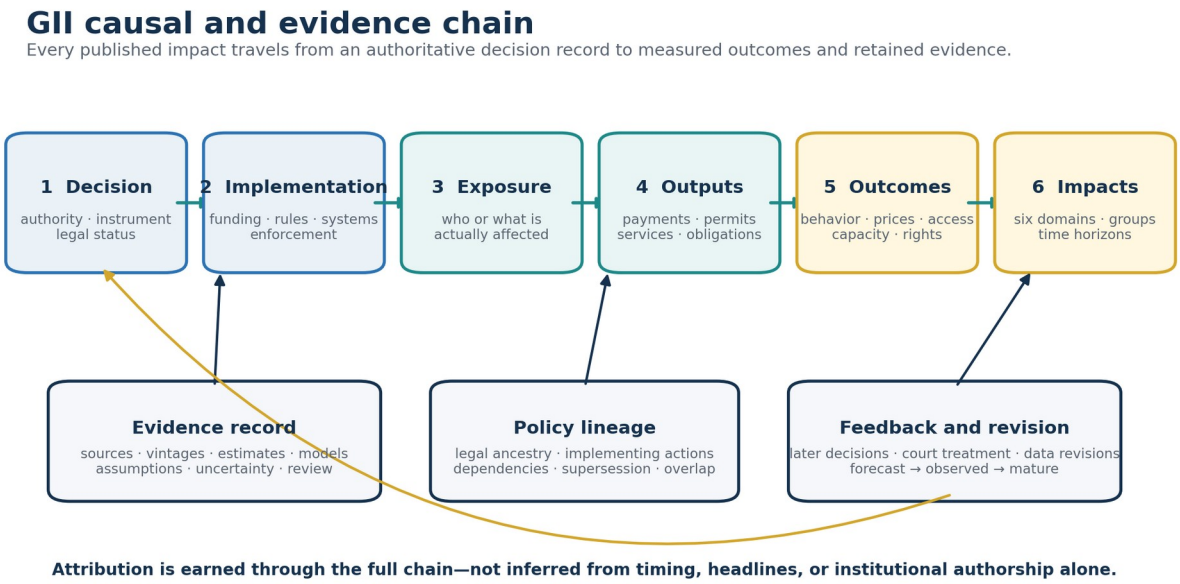


Figure 1. GII causal and evidence chain

The GII causal chain contains seven analytically distinct stages.

1. **Authority and decision.** An institution acts through an identifiable legal, fiscal, administrative, judicial, or monetary instrument.
2. **Implementation.** Responsible actors translate the decision into rules, funding, systems, enforcement, guidance, or service changes.
3. **Exposure.** People, entities, markets, or jurisdictions are actually subject to or able to use the change.
4. **Outputs.** Immediate deliverables occur, such as payments, permits, inspections, notices, rate changes, or revised legal rules.

5. **Intermediate outcomes.** Behavior, prices, eligibility, investment, service use, compliance, or institutional capacity changes.
6. **Longer-run impacts.** Welfare, health, rights access, productivity, public capacity, business structure, trade, or fiscal conditions change.
7. **Feedback and revision.** Courts, legislatures, agencies, markets, public entities, and affected groups adapt; source data and estimates are revised.

Temporal order is necessary but not sufficient for causality. Each arrow is a hypothesis that requires evidence. An implementation failure can break the chain; a judicial stay can pause it; a later statute can supersede it; and an external shock can alter the outcome without changing the decision.

4.1 Policy lineages

The policy lineage is the unit that connects related decision events. A lineage may include a statute, appropriations, a final rule, state implementation, litigation, a court ruling, revised guidance, and later reauthorization. Each event retains its institutional origin, while the lineage records relationships such as:

- authorized by;
- implements;
- funds or rescinds;
- modifies or supersedes;
- stays, enjoins, vacates, reverses, or affirms;
- delegates to;
- triggers implementation by;
- changes exposure to; and
- contributes evidence about an outcome.

Legal provenance and causal attribution are not the same edge. A rule may be legally authorized by a statute, but the incremental outcome change may be primarily associated with a later funding decision. A court may interpret the statute without being the sole empirical cause of every downstream condition. GII publishes both the legal chain and the estimated causal contribution.

4.2 Forecast, observed, and revised states

Every impact estimate is labeled as one of the following:

- **Forecast:** expected effect before sufficient post-exposure evidence;

- **Nowcast:** early estimate using incomplete or high-frequency information;
- **Observed—provisional:** post-exposure causal estimate subject to revision;
- **Observed—mature:** estimate using a declared outcome window and stable data;
- **Revised:** updated because of new data, model correction, legal change, or improved design;
- **Withdrawn:** no longer supported, with the prior record preserved.

Forecasts and observations may be compared for calibration, but they are never averaged into one effect. Proper scoring rules and forecast-accuracy research provide tools for evaluating whether probabilistic forecasts were calibrated once outcomes mature (Gneiting & Raftery, 2007; Hyndman & Koehler, 2006).

5. Unit of Analysis and Decision Impact Record

5.1 The decision event

The atomic unit is the Government Decision Event (GDE): one authoritative action by one primary decision origin at a defined decision time. If one document contains separable provisions with different populations, effective dates, or causal pathways, it may produce multiple provision-level events linked to the parent document.

Each event receives an identifier in the form:

GII-US-[YEAR]-[ORIGIN]-[TYPE]-[SERIAL]-[VERSION]

The identifier is stable across data revisions. A new legal action receives a new event identifier; a corrected estimate receives a new version of the existing event record.

5.2 Minimum Decision Impact Record

Record block	Required fields
Identity	GII ID, official title, citation, document number, docket or bill number, source URL, content hash
Origin	Branch, institution, component, decision maker, instrument type, authority
Lifecycle	Announcement, decision, publication, effective, implementation, enforcement, stay, sunset, repeal, supersession dates
Legal status	Proposed, final, effective, stayed, enjoined, vacated, reversed, superseded, expired; jurisdiction and precedential reach

Record block	Required fields
Classification	Policy topic, impact domains, geography, industries, programs, affected populations and entities
Causal model	Theory of change, exposure definition, counterfactual, estimand, analysis window, confounders
Evidence	Datasets, vintages, methods, assumptions, code version, robustness tests, peer or analyst review
Results	Raw estimates, uncertainty intervals, magnitude, direction, confidence, coverage, distribution, maturity
Lineage	Parent/child events, implementing actions, legal relationships, overlap and supersession rules
Governance	Analyst, reviewer, conflicts, publication date, revision reason, correction history

5.3 Multiple clocks

A government decision can have an announcement date, enactment or decision date, publication date, effective date, implementation start, enforcement date, litigation date, and sunset. GII stores each clock. The causal event time is selected for each outcome based on actual exposure, not automatically the date of a press release or signature.

Anticipation windows are required where firms, households, markets, or public agencies can respond before legal effectiveness. Judicial events additionally store mandate issuance, stays, remands, subsequent lower-court action, and compliance dates where relevant.

6. Domain Measurement Framework

6.1 People and households

The people-and-households domain measures direct and indirect changes in lived conditions. Its subpillars are:

- material resources: disposable income, taxes, transfers, prices, debt, wealth, housing and energy costs;
- work: employment, hours, wages, benefits, job transitions, labor-force participation;
- health and safety: morbidity, mortality risk, insurance, care access, injury, environmental exposure, food and product safety;

- rights and practical access: legally recognized coverage, voting or procedural access, due process, privacy, mobility, expression, nondiscrimination, and actual ability to exercise a right or obtain a service;
- public-service experience: eligibility, take-up, processing time, denial, error, availability, and service quality; and
- administrative burden: learning, compliance, documentation, waiting, travel, and reporting time.

Rights indicators are kept distinct from approval or ideology. A court's prevailing party, the political identity of an institution, or an analyst's view of a constitutional issue is not an outcome orientation. GII records measurable expansion or contraction of defined legal coverage and practical access, along with burdens and uncertainty, while preserving a narrative legal analysis for consequences that cannot be validly scored.

6.2 United States economy

The economy domain includes real output and income, employment and wages, prices, productivity, investment, financial conditions, business formation, and federal fiscal effects. National aggregates are decomposed by region and industry when exposure varies.

Several safeguards are necessary. GDP is an important production measure but not a full welfare measure. Inflation is evaluated relative to a declared target band or policy context rather than coded as "lower is always better." A deficit change is reported as a fiscal effect and not automatically coded as a benefit or burden. Monetary-policy effects are modeled with lags and multiple transmission channels. Economic data vintages are retained because official estimates are revised; BEA, for example, publishes advance, second, third, annual, and comprehensive GDP estimates (BEA, 2026a, 2026b).

6.3 Public entities

This domain covers federal implementing organizations and affected state, tribal, territorial, and local governments, as well as public schools, universities, hospitals, transit systems, utilities, and other publicly controlled entities. Subpillars include:

- own-source revenue, intergovernmental transfers, spending, debt, and capital needs;
- funded and unfunded mandates;
- workforce, systems, procurement, and administrative capacity;
- service volume, quality, access, and timeliness;
- legal exposure, litigation, compliance, and reporting burden; and
- resilience, emergency response, and continuity.

Federal transfers are not counted as simultaneous net benefits to the recipient and net costs to the federal government without reconciliation. GII distinguishes the resource transfer, financing source, administrative cost, behavioral response, and service outcome.

6.4 Nonprofits and civil society

Nonprofit impact is not reducible to federal grants. Relevant measures include grants and contracts, donations, program-service revenue, workforce, volunteer activity, demand for services, compliance burden, program capacity, organizational entry and exit, and access to affected communities. IRS Form 990 data support organization-level and sector-level analysis but arrive with filing lags and do not cover every small exempt organization with equivalent detail. USAspending supports federal award exposure, while household and service datasets can measure demand or outcomes.

Subgroup reporting distinguishes charities, foundations, associations, religious organizations where data permit, advocacy and civic organizations, educational and health nonprofits, and other exempt entities. GII does not infer mission success from revenue growth alone.

6.5 International trade

The trade domain measures imports, exports, tariff rates and incidence, customs revenue, trade balance components, prices, quantities, supply-chain reorientation, foreign direct investment, retaliation, export controls, and partner-country exposure. Census international trade data and USITC DataWeb provide official product, geography, and tariff information (U.S. Census Bureau, 2026b; USITC, 2026).

More exports are not automatically a net benefit, and fewer imports are not automatically a benefit. Consumer prices, producer surplus, input costs, government revenue, exchange rates, productivity, strategic resilience, and retaliation can move in different directions. GII therefore reports stakeholder incidence and avoids using the trade balance as a standalone welfare score.

6.6 Private enterprise

The private-enterprise domain measures revenue, cost, employment, wages, compliance and reporting burden, taxes and subsidies, capital access, investment, productivity, innovation, market entry and exit, concentration, competition, and supply-chain exposure. Results are decomposed by industry, geography, firm size, ownership where appropriate, and public versus privately held status.

SEC EDGAR supports public-company filings and structured financial facts, while BLS, Census, BEA, and agency data support sector and establishment outcomes. Private-company results will be less complete

without licensed sources; GII reports that coverage gap rather than allowing public-company visibility to stand in for all enterprise.

6.7 Cross-cutting distributional overlays

Every applicable domain is analyzed across relevant groups. Possible overlays include income and wealth, age, race and ethnicity under applicable federal standards, sex, disability, veteran status, household composition, geography, rurality, tribal and territorial status, industry, occupation, firm size, nonprofit type, trade partner, and public-entity type.

Distribution is not an afterthought to a national mean. Benefits and burdens are tabulated separately. Small cells, confidential data, and sensitive attributes are protected through aggregation, suppression, and approved secure-data methods. GII does not create individual political, legal, or socioeconomic profiles.

7. Causal Identification and Evidence Grading

7.1 Counterfactual specification

Every scoreable outcome requires:

- a treatment or exposure definition;
- the affected unit and population;
- a counterfactual or comparison strategy;
- a preregistered estimand;
- pre- and post-periods suited to the mechanism;
- known concurrent policies and shocks;
- tests of pretrends or identifying assumptions;
- uncertainty intervals; and
- robustness and falsification checks.

Policy endogeneity is a default concern: governments often act because outcomes are already changing. A rise in unemployment after an emergency program does not show that the program caused unemployment; the program may have responded to the same shock. Reverse causality and selection into exposure must be addressed directly. Each preregistered estimand states its target quantity in plain language as well as notation, following the discipline described by Lundberg, Johnson, and Stewart (2021).

7.2 Evidence ladder

The operational whitepaper's six-level A–F ladder is the normative evidence scale for GII publications. The five conceptual tiers below group that ladder for research exposition; research tiers A–E correspond to operational levels A/B, B/C, C/D, D/E, and E/F respectively.

Tier	Evidence design	Appropriate interpretation
A	Randomized, quasi-random, strong natural experiment, validated RD, or similarly persuasive identification	High-confidence causal estimate within the identified population and assumptions
B	Modern DiD/event study, synthetic control, strong IV, comparative interrupted time series, or well-validated structural design	Causal estimate with material assumptions and required robustness tests
C	Single-series intervention model, calibrated microsimulation, input-output/CGE model, matched observational design, or triangulated panel model	Model-based or moderately identified contribution estimate
D	Descriptive outcome change with mechanism evidence, implementation evaluation, expert-coded legal or institutional assessment	Association or contribution; not a strong standalone causal claim
E	Forecast, scenario, qualitative hypothesis, unverified early signal, or incomplete data	Exploratory; no mature observed-impact claim

Evidence tier and numerical confidence are related but not identical. A nominally strong design can have unreliable data, implementation failure, weak external validity, or undisclosed analytical flexibility. Conversely, multiple weaker sources can converge on a credible bounded conclusion.

7.3 Method assignment

- **Staggered state or local implementation:** cohort-specific difference-in-differences with dynamic effects and appropriate comparison groups.
- **Single national intervention:** synthetic control, comparative time series, structural model, or interrupted time series with explicit concurrent-event controls.
- **Eligibility or funding threshold:** regression discontinuity or randomized encouragement where valid.
- **Judicial decision with differential legal exposure:** state/circuit exposure designs, case-lineage analysis, event studies, and implementation measurement.

- **Tax, transfer, or eligibility change:** administrative microdata or microsimulation paired with observed take-up and behavioral evidence.
- **Trade measure:** product-country exposure, customs/tariff variation, firm or industry panels, pass-through and retaliation analysis.
- **Monetary decision:** high-frequency identification for immediate financial response plus lagged macro models for realized outcomes.
- **Company-market response:** short-window event study for repricing, followed by filings, investment, employment, and operating outcomes.
- **Rights or institutional effect without valid scalar outcome:** legal mapping, structured process tracing, coverage/access measures, and qualitative impact statement.

7.4 Attribution versus responsibility

GII estimates causal contribution; it does not adjudicate moral blame, political credit, legal liability, or institutional motive. Branch and actor labels describe where decisions were made and implemented. Lineage contribution weights, when used, are sensitivity-tested and never presented as a definitive allocation of political responsibility.

8. Index Construction

8.1 Scorecard, not a hidden scalar

The production GII is a decomposable scorecard with four primary quantitative dimensions and two mandatory qualifiers:

Measure	Scale	Meaning
GII-M: Impact Magnitude	0–100	Material size, reach, persistence, and systemic spillover of attributable impact
GII-D: Directional Balance	−100 to +100	Balance of preregistered measurable benefits and burdens; not ideology or approval
GII-C: Attribution Confidence	0–100	Strength of causal identification, data, mechanism evidence, robustness, and transparency
GII-K: Distributional Concentration	0–100	Concentration of benefits or burdens relative to the applicable reference population
GII-V: Coverage	0–100%	Share of applicable, preregistered outcome weight supported by usable evidence

Measure	Scale	Meaning
GII-S: Maturity	Categorical	Forecast, nowcast, provisional observed, mature observed, revised, or withdrawn

The primary "GII" value is GII-M because the project's central question is the magnitude of government-attributable impact. Direction, confidence, concentration, coverage, and maturity must appear beside it. A GII-M of 70 does not mean a decision was 70 percent good, popular, constitutional, efficient, or successful.

8.2 Indicator-level magnitude

For each outcome indicator, four components are scored on domain-specific, published anchors and later calibrated against a historical corpus:

- **E — Effect severity:** absolute standardized effect per affected unit or other declared materiality measure;
- **R — Reach:** proportion and count of the relevant people, entities, geography, activity, or economic base exposed;
- **P — Persistence:** measured or supported duration, reversibility, and lag structure; and
- **S — Systemic spillover:** demonstrated propagation across institutions, domains, industries, or jurisdictions.

The provisional v1.0.0 weighting rule, retained unchanged in v1.1.1, is:

$$M_i = 0.40E_i + 0.25R_i + 0.20P_i + 0.15S_i$$

These weights are design priors, not empirical truths. Production release requires sensitivity analysis using equal weights, expert-derived alternatives, and empirically calibrated variants. If a source estimates total rather than per-unit effect, reach is not counted again without reconciliation. The raw estimate, denominator, unit, and confidence interval remain primary data.

Implementation exposure is represented by u_i, from 0 to 1. Potential magnitude and realized magnitude are both retained:

$$M_i,\text{realized} = u_i \times M_i,\text{potential}$$

Proposed or blocked actions may have forecast potential magnitude but zero observed exposure.

8.3 Directional balance

Each indicator receives a preregistered orientation:

- +1 for an outcome direction treated as a measurable benefit;
- −1 for a measurable burden;
- 0 for magnitude-only or genuinely ambiguous outcomes; or
- a target-band function for measures where neither simple increase nor decrease is uniformly beneficial.

For a set of oriented indicators:

$$D = 100 \times [\Sigma_i \in O (w_i \times o_i \times M_i)] / [\Sigma_i \in O (w_i \times M_i)]$$

A value near zero can mean offsetting large benefits and burdens; it does not necessarily mean little impact. GII-M resolves that ambiguity. Both sums run over the oriented-indicator set O — indicators with a preregistered, non-null orientation rule. Indicators scored 0 under a validated ambiguity rule remain in O; magnitude-only indicators are excluded, and directional coverage is published whenever indicators outside O carry material magnitude. Indicator orientation rules are published before outcome inspection where possible. Transfers, taxes, prices, trade flows, deficits, and court dispositions are not assigned a simplistic universal sign.

8.4 Attribution confidence

Attribution confidence is a weighted assessment of five documented components:

$$C = 0.30I + 0.20Q + 0.20F + 0.15B + 0.15T$$

where:

- **I** is identification strength;
- **Q** is data reliability, including accuracy, completeness, applicability, and vintage;
- **F** is implementation and mechanism fidelity;
- **B** is robustness, triangulation, placebos, negative controls, and independent consistency; and
- **T** is transparency, reproducibility, disclosure, and review.

GAO's data-reliability guidance supports a risk-based assessment of accuracy, completeness, and applicability rather than reliance on source reputation alone (GAO, 2020). Confidence bands are provisionally labeled High (80–100), Substantial (65–79), Moderate (45–64), Limited (25–44), and Exploratory (0–24). Numerical thresholds must be validated during the pilot.

8.5 Distributional concentration

Benefits and burdens receive separate concentration measures. For impact-share vector **p** and reference-population or entity-share vector **q**:

$$K = 50 \times \Sigma|p_g - q_g|$$

with shares expressed as fractions summing to one, producing a 0–100 total-variation measure (equivalently, $K = 100 \times \frac{1}{2} \times \Sigma|p_g - q_g|$). The choice of reference population is disclosed. Classified shares are renormalized before computation, and the unclassified impact share is published alongside K, flagging the score when it exceeds a provisional 20 percent threshold. A high K does not inherently indicate injustice; it indicates that impact is concentrated and requires group-level interpretation.

8.6 Coverage and missingness

Coverage is calculated over the preregistered set of applicable domains and outcomes. Inapplicable domains receive no weight; missing applicable outcomes remain missing. Weights are not silently redistributed. A headline event score is withheld when usable evidence covers less than 70 percent of applicable outcome weight or when no direct stakeholder outcome is available. Provisional research estimates below that threshold may be published only with an explicit incomplete-coverage label.

8.7 Event and domain aggregation

Indicators are grouped into correlated outcome families before aggregation so that multiple versions of the same construct do not dominate. Within each applicable domain, equal outcome-family weights are the default starting rule. Event-level magnitude is an equal-weight average across applicable domains unless a documented domain weighting has been approved before analysis. Alternative weight sets are included in sensitivity results.

An event cannot improve its score merely because more similar indicators are available. Coverage rewards breadth of evidence; magnitude reflects measured impact.

8.8 Lineage and double-counting rules

The same outcome cannot be fully attributed to a statute, rule, implementing grant, and court decision in the national total. GII applies four controls:

1. outcome estimates receive a lineage and exposure identifier;
2. parent and child events are marked as overlapping, incremental, superseding, or independent;

3. only non-overlapping incremental contributions enter aggregate pressure; and
4. alternative contribution allocations are sensitivity-tested.

Supreme Court results are separately displayed but included once in Judiciary. Transfers are reconciled across payers and recipients. Public-company market reactions are not added to later operating effects without addressing anticipation and overlap.

8.9 National, branch, and domain time series

For active event **e** at time **t**, define a non-overlapping evidence-weighted contribution:

$$P_{e,t} = a_{e,t} \times M_e \times (C_e / 100) \times n_e$$

where **a** is the observed exposure or distributed-lag share and **n** is the non-overlap factor. The monthly raw impact pressure is the sum of event contributions. Once a historical benchmark is available, raw pressure is converted to a 0–100 empirical percentile relative to a fixed reference window. The national directional balance is the magnitude-weighted direction of active contributions. Confidence enters aggregation only at this cross-event level, where unweighted summation would allow weakly identified events to move the national series as much as well-identified ones; event-level magnitude and confidence remain unmultiplied. The linear $C_e/100$ weighting is itself a provisional design choice: the pilot must test alternatives, including threshold inclusion (for example, $C_e \geq 45$) and concave weightings, and report series sensitivity to that choice.

Branch, institution, policy-topic, domain, geography, and stakeholder series are calculated by filtering the same event contributions. A fixed benchmark and vintage-preserving back series prevent each new release from silently redefining history.

8.10 Uncertainty and sensitivity

Each published aggregate includes:

- statistical uncertainty from effect estimates;
- score ranges under alternative normalization and weighting;
- sensitivity to event inclusion and overlap allocation;
- sensitivity to causal design and analysis window;
- source-vintage and revision effects; and
- an influence analysis showing which events and assumptions drive the result.

If reasonable specifications change the substantive conclusion, the index must say so.

9. Data Provenance, Reproducibility, and Governance

9.1 Source hierarchy

GII uses the following evidence hierarchy:

1. authoritative legal and government publications for decision facts;
2. official statistical and administrative data for outcomes;
3. independently reproducible peer-reviewed studies and high-quality public evaluations;
4. audited or methodologically documented third-party data;
5. licensed commercial sources where they materially close a known gap; and
6. news and public commentary for discovery only, not as standalone causal evidence.

An official source can still contain incomplete, revised, or purpose-limited data. Source authority does not eliminate reliability assessment.

9.2 Provenance model

Every value retains source URL, retrieval time, content hash, licensing or use note, dataset and table identifier, release date, vintage, transformation, code version, analyst decision, and reviewer action. W3C PROV-O provides a standard conceptual vocabulary for entities, activities, agents, and derivations, while DCAT supports interoperable dataset and data-service metadata (W3C, 2013, 2024). GII can implement these standards without requiring every consumer to use RDF.

9.3 Reproducibility

Published methods, code, source manifests, and synthetic or aggregate test data should allow an independent analyst to reproduce public results where law and licensing permit. Restricted microdata remain in approved environments; release artifacts contain disclosure-reviewed aggregates and sufficient documentation. Analysis of public and disclosure-protected aggregate data constitutes secondary analysis without human-subjects intervention; any restricted-microdata analysis will document the applicable institutional-review or data-steward determination. Outcome maps, orientation rules, and analysis plans are preregistered before outcome inspection through a public third-party registry (for example, OSF) and, in parallel, through GII's own publication registry using cryptographic content hashes and timestamps included in each release package. The National Academies distinguishes computational reproducibility from independent replicability and recommends transparent evidence and methods (NASEM, 2019). GII should pursue both.

9.4 Human and automated review

Automated extraction, entity resolution, document classification, and anomaly detection can reduce cost and improve timeliness. They cannot make final legal-scope, outcome-orientation, causal-design, or publication decisions without accountable human review. Any model-assisted field records the model or rule version, source text span, confidence, and reviewer disposition.

9.5 Independence and conflict controls

The index requires:

- a published methodology and change log;
- analyst and reviewer conflict disclosures;
- separation of sponsorship from scoring decisions;
- two-person review of legal coding and causal claims;
- external methods review for major revisions;
- a public challenge and correction channel;
- embargoed, scheduled releases; and
- preservation of prior versions and reasons for change.

First-party source disclosure. The Presidential Economic Pressure Index is developed by the author and published by API Automations, which supplies the live index platform for this research program. The relationship is disclosed as a potential conflict of interest and is mitigated structurally: PEPI values are excluded from all GII score computations, PEPI-derived candidates pass the same authoritative-record, causal, coverage, and review gates as any other external source, and adapter promotion decisions are recorded with reviewer identity. External methods review of the PEPI boundary controls is a Phase 2 requirement.

GII is not a government publication and must not imply endorsement by a court, agency, statistical office, or other public body.

10. Public-Data-First Research Architecture

The decision registry can be built largely from free official APIs, bulk downloads, feeds, and publications. Congress.gov and GovInfo supply legislative metadata and official documents; FederalRegister.gov, Regulations.gov, RegInfo, and eCFR supply regulatory and presidential materials; SupremeCourt.gov,

GovInfo, the federal courts, FJC, and CourtListener support judicial records; and the Federal Reserve publishes FOMC decisions and historical materials.

Outcome layers can begin with Census, BLS, BEA, Treasury Fiscal Data, USAspending, CBO, IRS Form 990, SEC EDGAR, Census international trade, USITC DataWeb, CDC, CMS, EPA, FEMA, EIA, and USDA. These sources differ in cadence, geography, history, revision policy, privacy, and API access. Their suitability must be assessed at the indicator level.

API access is preferred to scraping. Bulk files are preferred when complete historical snapshots or high-volume ingestion are needed. RSS, sitemaps, and structured HTML can fill gaps. HTML scraping is the last resort and must follow source terms, robots instructions, rate limits, and change monitoring. Source documents are snapshotted with hashes so later corrections or silent revisions can be detected.

10.1 PEPI as the initial external source adapter

The Presidential Economic Pressure Index (PEPI), published by API Automations at pepi.apiautomations.com, is a useful first adapter because it already organizes presidential economic-policy events into stable records, lifecycle states, source links, forecast and outcome fields, confidence flags, and versioned methodology. The integration baseline reviewed for this revision was PEPI methodology v1.2.13 and schema v1.3.6 (API Automations, 2026a–e). GII should record the PEPI methodology and schema versions returned or disclosed at retrieval rather than assume those versions remain current.

PEPI and GII answer different questions. PEPI measures the degree to which presidential economic-policy pressure is implemented, diluted, delayed, reframed, reversed, or abandoned. GII measures attributable impact magnitude, direction, distribution, confidence, coverage, and evidence maturity across multiple government institutions and stakeholder domains. Therefore:

- a PEPI composite or reversal score is never copied into **GII-M**, **GII-D**, or **GII-C**;
- PEPI records may nominate candidate decisions, provide source provenance, identify implementation milestones, supply trade or price observations, and help select pilot cases;
- a PEPI event becomes a scoreable GII event only after the GII legal-effect, implementation, exposure, counterfactual, coverage, and review gates are satisfied; and
- PEPI confidence describes PEPI classification or resolution certainty, while **GII-C** describes confidence in causal attribution. The two fields must retain distinct names and definitions.

PEPI-to-GII field crosswalk

PEPI element	GII destination	Integration rule
event_id (PEPI-native identifier)	external_source_id	Preserve exactly and pair with a separately assigned GII stable ID
domain	policy_topic / source tag	Do not treat PEPI's topic domain as one of GII's six stakeholder impact domains
status	candidate-event lifecycle	Map detected or confirmed to watch/forecast review; monitoring to provisional monitoring; resolved to resolution review; rejected to screened out; never infer observed impact from status alone
threat_date, threat_source	announcement clock and provenance	Retain as a candidate decision date and source; do not treat an announcement as legal effect
statutory_authority	authority and legal-lineage fields	Verify against authoritative legal text before publication
stated_deadline	implementation milestone	Use as an expected clock; distinguish it from effective and exposure dates
outcome_date, outcome_type	implementation/resolution evidence	Store the source classification and independently verify the operative action
confidence	source_classification_confidence	Never map to GII causal confidence
composite_score, reversal_magnitude	contextual follow-through attributes	Retain for source context or sensitivity analysis; exclude from GII scoring inputs
tariff, CPI, passthrough and consumer-burden fields	candidate outcome observations	Reconcile units, vintages, baselines, geography, incidence, and causal design before use
sector and market-impact fields	exposure or outcome candidates	Treat short-window market reactions as contextual event-study evidence, not total social impact
Federal Register documents and source links	decision documents and provenance edges	Preserve authoritative URI, retrieval time, checksum, and relationship to the event

PEPI's indeterminate and pending outcome states are missing or unresolved states, not numerical zeroes. They must not enter aggregation as full implementation, reversal, no effect, or neutral direction. Near-duplicate events collapsed by PEPI should still be checked against the GII Policy Lineage Graph because the two systems can define duplication at different levels. External identifiers and field names are stored verbatim only within the adapter layer, as source data; they are not surfaced in GII records, and their form carries no GII classification, endorsement, or directional meaning.

Retrieval, security, and freshness controls

The adapter should use https://api.pepi.apiautomations.com and send the premium credential in the X-API-Key header. The credential must be provided to the runtime through a secret manager or a PEPI_API_KEY environment variable; it must not be embedded in source code, documents, notebooks, request logs, screenshots, or published reproducibility packages. A premium key should never be pasted into a public issue or conversational transcript.

Initial discovery and health checks can use PEPI's index, statistics, snapshot, source-status, and health resources. Event-level ingestion should use the event list and event detail resources, with related Federal Register, timeline, sector-impact, tariff, indicator, research-source, and court records as applicable. Premium economic, price, market, tariff, and trade resources may supplement candidate outcome fields, subject to the same GII source-quality and causal-review standards as any other upstream dataset.

Every retrieval log should include the request timestamp, endpoint, nonsecret parameters, HTTP result, payload checksum, source generated_at value, API path version, disclosed PEPI methodology and schema versions, pagination state, and transformation version. The ingest must stop or quarantine records when freshness, schema, pagination, or provenance checks fail. Source-status metadata are monitoring evidence, not a guarantee of currentness; the adapter must compare timestamps and expected cadences before use.

The source adapter is intentionally modular. If PEPI changes its methodology or schema, a new adapter mapping and compatibility test can be released without retroactively rewriting GII estimates. Reprocessing must create a new GII data vintage with a documented diff.

11. Validation Program and Research Propositions

11.1 Pilot design

The first empirical program should assemble a stratified historical benchmark of approximately 24–36 federal decision lineages from 2015–2025. Selection should occur before scores are calculated and cover:

- multiple presidential administrations and Congresses;
- legislative, presidential, agency, independent-regulatory, judicial, Supreme Court, and monetary origins;
- regulatory and deregulatory actions;

- spending increases, reductions, taxes, and transfers;
- nationwide and staggered implementation;
- multiple policy topics and all six impact domains;
- decisions expected to create both benefits and burdens; and
- strong, moderate, and weak data environments.

Selection should not be based on whether a result supports a political narrative. A holdout set should test the stability of anchors, weights, and confidence grading.

11.2 Research propositions

1. **Lineage proposition:** Cross-institution policy lineages explain observed effects more accurately than isolated document attribution.
2. **Heterogeneity proposition:** National means materially understate variation across households, geographies, public bodies, nonprofits, industries, and firm sizes.
3. **Forecast-calibration proposition:** The divergence between forecast and observed impact varies systematically by decision type, domain, and evidence maturity.
4. **Implementation proposition:** Legal effectiveness alone is an incomplete proxy for realized exposure and impact.
5. **Measurement proposition:** A magnitude-direction-confidence scorecard is more stable and interpretable than a single positive/negative composite.
6. **Revision proposition:** Data vintages and legal-status changes materially affect a nontrivial share of early estimates.

11.3 Validation tests

- inter-rater reliability for legal and domain coding;
- backtesting against established evaluations;
- placebo and negative-control outcome tests;
- alternative counterfactual and event-window analyses;
- weight and normalization sensitivity;
- forecast calibration and accuracy;
- coverage-bias analysis;
- replication by an independent analyst; and
- adversarial review from methodologists with different policy perspectives.

Reviewer agreement is assessed with named statistics rather than an undefined standard: Krippendorff's alpha (interval) for the E, R, P, and S magnitude components and the I, Q, F, B, and T confidence components; weighted kappa or ordinal alpha for evidence-level and maturity assignment; and ICC(2,k) for composite magnitude. Provisional acceptance thresholds are $\alpha \geq 0.67$ for pilot learning and $\alpha \geq 0.80$ for production scoring, calibrated in the pilot.

12. Limitations and Failure Modes

GII cannot eliminate the fundamental difficulties of public-policy evaluation. Its principal limitations are:

- **No perfect counterfactual.** National decisions often lack an untreated comparison.
- **Policy endogeneity.** Government action responds to the conditions later used as outcomes.
- **Interference and spillovers.** Treated and comparison units affect one another.
- **Concurrent decisions.** Multiple branches and jurisdictions act during the same window.
- **Long and unequal lags.** Legal, health, investment, fiscal, and institutional effects mature at different speeds.
- **Anticipation and adaptation.** Behavior changes before effective dates and continues after repeal.
- **Measurability bias.** Economic and administrative outcomes may be easier to quantify than rights, dignity, trust, or institutional legitimacy.
- **Private-data gaps.** Privately held firms and small nonprofits have less timely public reporting.
- **Judicial implementation.** A holding's real effect depends on mandate, remand, compliance, enforcement, and subsequent litigation.
- **Source revision.** Official economic, fiscal, and administrative data change after first release.
- **Weighting judgment.** Any composite reflects choices that require sensitivity analysis and public disclosure.
- **Direction contestability.** Some outcomes have no universal benefit/burden orientation.
- **Responsibility ambiguity.** Causal contribution is not equivalent to political, legal, or moral responsibility.

GII must not be used as a standalone measure of constitutionality, legality, public approval, partisan performance, institutional legitimacy, or moral worth. It must not infer that unmeasured impact is zero. It must not compare branch scores without accounting for jurisdiction, decision volume, policy baseline, applicable domains, and evidence coverage.

13. Development Roadmap

Phase 1 — Foundation and registry

Operationalize the v1.1.0 taxonomy, Decision Impact Record, provenance fields, outcome dictionary, orientation rules, source registry, and review handbook. Build and validate the PEPI adapter as the first external-source walking skeleton, followed by ingestion for core legislative, regulatory, judicial, spending, statistical, and monetary sources.

Phase 2 — Historical pilot

Construct the stratified 2015–2025 lineage sample, estimate outcome effects, assess inter-rater reliability, calibrate component anchors, and document failure cases. Do not publish a national headline score during calibration.

Phase 3 — Beta index

Release event records, domain and institution dashboards, methodology code, revision histories, confidence and coverage, and a limited monthly beta. Publish sensitivity ranges and an external methods review.

Phase 4 — Production and API

Expand decision coverage, formalize a fixed benchmark window, publish stable API contracts, add bulk downloads and data citations, and establish scheduled monthly, quarterly, and annual releases.

Phase 5 — Expansion

Evaluate state and local subindices, international comparisons, licensed private-company data, richer nonprofit outcomes, causal-text and legal-provision tooling, and forecast calibration products. Methodology changes that break historical comparability require a major version.

14. Conclusion

The Government Impact Index begins from a simple observation: government output is not the same as government impact. A statute, rule, order, appropriation, court decision, or rate action matters because of the implementation and outcomes that follow—not merely because the action appears in an official record.

Measuring those outcomes requires institutional literacy, legal provenance, causal inference, distributional analysis, transparent aggregation, and revision discipline. It also requires restraint. Some effects can be monetized; others cannot. Some decisions create measurable gains and burdens at once. Some conclusions will remain uncertain even after substantial analysis.

GII is therefore designed as an evidence system before it is a ranking. Its primary score measures attributable magnitude. Its directional balance shows the measured mix of benefits and burdens. Confidence, coverage, maturity, and distribution reveal what the headline cannot. Policy lineages preserve the reality that Congress, presidents, agencies, courts, implementers, and the Federal Reserve can jointly shape one outcome.

If built according to these principles, GII can provide a durable public answer to a question that is routinely debated but rarely measured in a unified way: what did a government decision change, who experienced the change, how large and persistent was it, and how strong is the evidence that the decision caused or materially contributed to the result?

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

API Automations. (2026a). *Presidential Economic Pressure Index methodology*. https://pepi.apiautomations.com/methodology/

API Automations. (2026b). *PEPI API documentation*. https://pepi.apiautomations.com/api/

API Automations. (2026c). *PEPI data sources and provenance*. https://pepi.apiautomations.com/data-sources/

API Automations. (2026d). *PEPI data dictionary*. https://pepi.apiautomations.com/data-dictionary/

API Automations. (2026e). *PEPI changelog*. https://pepi.apiautomations.com/changelog/

Arrow, K. J., Cropper, M. L., Eads, G. C., Hahn, R. W., Lave, L. B., Noll, R. G., Portney, P. R., Russell, M., Schmalensee, R., Smith, V. K., & Stavins, R. N. (1996). Is there a role for benefit-cost analysis in environmental, health, and safety regulation? *Science*, *272*(5259), 221–222. https://doi.org/10.1126/science.272.5259.221

Bureau of Economic Analysis. (2026a). GDP revision information. https://www.bea.gov/gdp-revision-information

Bureau of Economic Analysis. (2026b). GDP release—Additional information. https://www.bea.gov/news/gdp-release-additional-information

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Centers for Disease Control and Prevention. (2024). Legal epidemiology. https://www.cdc.gov/phlp/php/about/legal-epidemiology.html

Comparative Agendas Project. (2026). CAP topics: Master codebook. https://www.comparativeagendas.net/pages/master-codebook

Congressional Budget Office. (2022). *Current work on the distributional analysis of household income resulting from policy changes* (Working Paper 2022-09). https://www.cbo.gov/publication/58508

Congressional Budget Office. (2023). *CBO describes its cost-estimating process*. https://www.cbo.gov/publication/59003

Congressional Budget Office. (2026a). Cost estimates. https://www.cbo.gov/cost-estimates

Congressional Budget Office. (2026b). Distributional analysis. https://www.cbo.gov/about/distributional-analysis

Congressional Research Service. (2024). *Constitution of the United States of America: Analysis and interpretation—2024 supplement*. U.S. Government Publishing Office. https://constitution.congress.gov/static/files/GPO-CONAN-2024-SUPP.pdf

Federal Judicial Center. (2026). Federal court cases: FJC Integrated Database, 1979 to present. https://www.fjc.gov/research/federal-court-cases-fjc-integrated-database-1979-present

Free Law Project. (2026). CourtListener REST API, version 4.4. https://wiki.free.law/c/courtlistener/help/api/rest/v4/overview

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

Library of Congress. (2026). Congress.gov API. https://www.loc.gov/apis/additional-apis/congress-dot-gov-api/

Lopez Bernal, J., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, *46*(1), 348–355. https://doi.org/10.1093/ije/dyw098

Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. *American Sociological Review*, *86*(3), 532–565. https://doi.org/10.1177/00031224211004187

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, *35*(1), 13–39. https://www.jstor.org/stable/2729691

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. National Academies Press. https://doi.org/10.17226/25303

OECD, European Union, & European Commission Joint Research Centre. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing. https://doi.org/10.1787/9789264043466-en

Office of Management and Budget. (2003/2025). *Circular A-4: Regulatory analysis* (September 17, 2003; reinstated February 12, 2025). https://www.whitehouse.gov/wp-content/uploads/2025/08/CircularA-4.pdf

Office of Management and Budget. (2020). *M-20-12: Phase 4 implementation of the Foundations for Evidence-Based Policymaking Act of 2018—Program evaluation standards and practices*. https://www.whitehouse.gov/wp-content/uploads/2020/03/m-20-12.pdf

Office of Management and Budget. (2025). *M-25-15: Rescission and reinstatement of Circular A-4*. https://www.whitehouse.gov/wp-content/uploads/2025/03/M-25-15-Recission-and-Reinstatement-of-Circular-A-4.pdf

Public Law 115-435. (2019). *Foundations for Evidence-Based Policymaking Act of 2018*. https://www.govinfo.gov/app/details/PLAW-115publ435

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A*, *168*(2), 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

Supreme Court Database. (2026). Online codebook and documentation. https://scdb.wustl.edu/documentation.php

Supreme Court of the United States. (2026). Opinions of the Court. https://www.supremecourt.gov/opinions/opinions.aspx

U.S. Census Bureau. (2026a). Census Data API: Available APIs. https://www.census.gov/data/developers/data-sets.html

U.S. Census Bureau. (2026b). International trade data API user guide. https://www.census.gov/foreign-trade/reference/guides/Guide_to_International_Trade_Datasets.pdf

U.S. Courts. (2026). Caseload statistics data tables. https://www.uscourts.gov/statistics-reports/caseload-statistics-data-tables

U.S. Government Accountability Office. (2020). *Assessing data reliability* (GAO-20-283G). https://www.gao.gov/products/gao-20-283g

U.S. Government Accountability Office. (2021). *Program evaluation: Key terms and concepts* (GAO-21-404SP). https://www.gao.gov/products/gao-21-404sp

U.S. Government Accountability Office. (2023). *Evidence-based policymaking: Practices to help manage and assess the results of federal efforts* (GAO-23-105460). https://www.gao.gov/products/gao-23-105460

U.S. Government Publishing Office. (2026a). United States Reports. https://www.govinfo.gov/help/usreports

U.S. International Trade Commission. (2026). DataWeb: U.S. trade and tariff data. https://dataweb.usitc.gov/

World Bank. (2026). Worldwide Governance Indicators. https://www.worldbank.org/en/publication/worldwide-governance-indicators

World Wide Web Consortium. (2013). PROV-O: The PROV ontology. https://www.w3.org/TR/prov-o/

World Wide Web Consortium. (2024). Data Catalog Vocabulary (DCAT), version 3. https://www.w3.org/TR/vocab-dcat-3/

Document control

Title: Government Impact Index (GII): A Causal, Multi-Institution Framework for Tracing the Effects of United States Government Decisions
Document type: Research Paper — Methodology Foundation
Version: 1.1.1
Date: July 19, 2026
Author: Kurt Sather
Research program: Government Impact Index (GII-US)
Concept credit: Randall Hall, CEO, BookedByFriday
Contributing ideation: Shivani D. Patel, CTO, API Automations
Status: Foundational working paper with PEPI source-adapter baseline for continued research and index development
License: Proprietary and confidential — internal use only. All rights reserved.
Distribution: Confidential — internal use only; not for public release. Prior versions are retained immutably as internal version references.
Cite as: Sather, K. (2026). Government Impact Index (GII): A Causal, Multi-Institution Framework for Tracing the Effects of United States Government Decisions (Version 1.1.1)