**Global Work-Injury Policy Dataset: Laws, Coverage and Replacement Rates in 188 Countries since the Industrial Revolution**

**Nate Breznau,** [breznau.nate@gmail.com](mailto:breznau.nate@gmail.com) **,** [ORCID](https://orcid.org/0000-0003-4983-3137)

German Institute for Adult Education – Leibniz Center for Lifelong Learning

**Abstract** (≈150 words)

This Data Descriptor introduces the *Global Work‑Injury Policy* (GWIP) *Longitudinal* dataset, an open resource tracing the introduction and evolution of work‑injury policies across 188 countries from 1838 to 2020. In any country‑year with evidence of a major legislation or outlay change, the file records the name of the governing law, the type of programme, and provides estimates of coverage and wage‑replacement generosity for both permanent and temporary work-related injury or illness. From the 2000s onward, additional data points are added even in the absence of policy changes – taking advantage of greater source material availability and increasing analytical power. The data can be combined with the GWIP Cross-Sectional, a previously published dataset that allows some harmonization with other resources and timing of first laws by occupational groups. Together these files enable researchers to examine the diffusion and development of workers’ compensation systems and engage in diverse research activities in welfare state, social policy and socio-economics. In the absence of evidence of a policy change or policy collapse (as during wars) I assume that policies do not change in between data points and provide code to fill in these data points and merge the data with other datasets. A detailed codebook and versioned datasets are available on the Harvard Dataverse. Replication and recommended usage scripts are available in the Online Repository (GitHub).

# Background & Summary

Workers’ compensation is one of the first forms of social welfare policy in human history 1. Its rise coincides with the Industrial Revolution because of the horrific injuries and deaths that occurred as a result the mechanization of extraction and production. The idea of insurance against risks associated with work had it origins prior to industrialisation. Early legal codes in Mesopotamia, Rome and China compensated certain injuries with, most often for soldiers. In the nineteenth century industrialisation increased workplace risk and social tensions. Prussian Chancellor Otto von Bismarck introduced a work-injury insurance scheme in 1884, creating the first modern social insurance programme for industrial injuries. Within a few decades, similar legislation spread through Europe and North America, although implementation varied greatly by country 2,3.

Despite its global diffusion, comparative data on work‑injury policies outside of the Global North or prior to the year 2000 was scarce or non-existent until now 4. The *Global Work‑Injury Policy Database* (GWIP) is a first attempt to fill these scientific data gaps. An earlier cross‑sectional release (GWIP\_v3) harmonised information on the first adoption of work‑injury laws, program types, coverage of various occupational groups and generosity of benefits for 189 countries as of 2020. The GWIP\_v3 data are available as file gwip\_v3.tab on the Harvard Dataverse and contain 34 variables for 189 countries (or territories in a few cases). It provides data on early adoption and comparative coverage and replacement rates in 2020, but the longitudinal perspective is missing.

The GWIP\_long v1.0 file finally adds the longitudinal development of coverage as a percentage of the labour force and replacement rate as a percentage of previous wages since the introduction of work-injury policy in each of 188 countries. It records 1,467 country‑year observations and includes further variables on qualifying periods and conditions, and notes on sources and assumptions. A value only exists in the data when primary or secondary sources confirm that a law was introduced or a policy changed. In the 2000s, and largely based on the regularity of Social Security Programs Throughout the World data, I was able to have observational data points every 4 years for most countries.

Between legislative changes, replacement rates are assumed to remain stable while coverage may change only when evidence suggests an increase. I do not code these assumptions into the data, but offer a workflow here to generate them. The resulting dataset enables analyses of policy diffusion, convergence and divergence, and facilitates linkage with other longitudinal social policy databases such as the Social Insurance Entitlements Dataset (SIED).

# Conceptual scope

The GWIP framework defines *work‑injury policy* as any statutory provision that compensates workers or their dependants for income loss arising from injury sustained in the course of employment. Programmes differ in type. There are employer‑liability schemes requiring firms to compensate injured workers directly, risk-pooling schemes with compulsory insurance managed by mutual funds or private carriers, and social insurance schemes that are publicly regulated and administered and often financed by payroll contributions. Work‑injury programmes often laid the groundwork for comprehensive national social security systems.

A work-injury policy is different from a disability policy. General disability is a form of social policy or insurance that supports those who are disabled, regardless of how they came to be disabled. Work-injury is specifically work-related. Its modern versions cover any injury or illness caused by work-related activities including commuting to work, working at a desk, working with dangerous chemicals, tripping and falling at work, and etc. with great variation by country and time period.

The GWIP Longitudinal is optimized for general cross-temporal and cross-spatial comparison. Therefore it trades of detailed accounts of the exact features of work-injury policies for the possibility to compare similarities across all countries with a population today of over 1 million, with several smaller countries included.

# Data sources

Data were compiled manually from primary legislation and authoritative secondary sources[[1]](#footnote-1). Primary sources were copies of the laws themselves, that were auto translated by Google Translate prior to 2023 and ChatGPT (various versions) after, when no translation was available. Secondary sources included scientific books and articles, the *Social Security Programs Throughout the World* (SSPW) series produced by the U.S. Social Security Administration, the International Labour Organization’s NATLEX database, and legislative bulletins from the U.S. Bureau of Labor. Each entry in the GWIP\_long references the sources used in a semi‑colon separated list and all sources are available in a public Zotero group[[2]](#footnote-2). The country identifiers follow the Correlates of War (COW) coding scheme (cow\_code), enabling linkage with other international datasets.

# Data extraction and coding

The cross-sectional GWIP has the year of the first work‑injury law (statute that explicitly compensates workers for occupational injury), the first insurance scheme (when risk pooling or social insurance replaced employer liability), the first dedicated fund and first social insurance enactment – extracted from the original and secondary sources. It also documents if the first law was carried over from colonial times or enacted new where applicable (labor\_workinjury\_firstnat\_carriedover). Program types were categorised based on financing and administration: employer liability, risk pooling, social insurance or national scheme. In the most recent release of the cross-sectional data, I added the first year of legal coverage for agricultural, blue-collar and white-collar workers to enable further investigation of the historical timings of laws by occupation.

For the cross-sectional GWIP v3, I extracted the coverage, replacement rates and benefit durations for permanent (1-year) and temporary (6-month) disabilities for the year 2020, or closest available. Replacement rates express the ratio of wage compensated, while durations reflect the maximum payment period (in months for permanent and weeks for temporary injuries). Where legislation specifies a lump sum rather than periodic benefits, this is noted. In this dataset the unit of analysis is the country and the data are in wide format.

For GWIP\_long, the unit of analysis is the country‑year and the data are in long format with countries nested in years. I identified each year in which a major legal change occurred or for which coverage and generosity estimates were available. Sometimes laws are recorded when this information is not available to nonetheless chart progression and changes to other aspects, for example transition from compulsory insurance to social insurance. Important values are the name of the statute in force, the social policy programme type, and a reference list. Replacement rates and durations are coded separately for permanent and temporary injuries. Coverage estimates represent the percentage of the labour force legally entitled to benefits and are sometimes derived from secondary sources and calculations based on historical labour force estimates. A qualifying period denotes the waiting period before benefits commence, and qualifying conditions specify occupational or employment requirements. Various notes variables provide additional context or caveats (e.g., whether benefits are lump sums or limited to specific sectors, or where certain atypical coding decisions had to be made).

# Primary variables of interest

Detailed definitions and further information about all variables in both the GWIP and GWIP\_long are available in the Codebook 6. Here I will go into detail only regarding two central variables. The centrepiece variables of the GWIP longitudinal are coverage (*labor\_workinjury\_coverage\_pct\_lf*) and wage replacement rate (*labor\_workinjury\_replacement\_rate\_perm* and *labor\_workinjury\_replacement\_rate\_temp*). In comparative welfare state research, these are the key variables used to identify the inclusivity and generosity of social policy to develop typologies, understand inequalities and engage in institutional analysis 7–10.

Coverage as a percentage of the labor force is measured as those who are legally covered out of the formal labor force. This excludes informal and shadow economy work, and subsistence activities like farming. Legal coverage comes at first from the legal text. If a law states that all “workers” are covered, the law usually defines what this means. In many countries a law covers all “workers” in establishments with a certain number of employees. This means that we need to estimate how much of the labor force works in such sized firms, often 5 or more. There are also laws in many countries that historically could exempt certain persons either from a national work-injury law, or from the definition “worker”. Many countries introduced export processing zones (EPZs) starting the 1960s and especially 1970s. These zones were legally exempt from most labor laws to encourage high productivity and high profits, to keep a country competitive globally. In these cases, I used available resources to estimate what percentage of the formal labor force likely worked in EPZs and subtracted these from the coverage.

Coverage comes with an anchor, as the ILO engaged in a massive undertaking to measure coverage in 2014 and then again in 20204,11. This provides an anchor measure for assistance in calculating other years that may contain missing values. Except for war and EPZs, coverage tends to remain steady or increase over time. In case there was much uncertainty in the sources, coverage was coded conservatively. The variable *coverage\_notes* and the Codebook provide additional details to understand some of the decisions and complexities in estimating coverage.

One issue with coverage is that there are cases where there is a known law, but I am unable to estimate coverage because there is no information about the law in any primary or secondary sources. For these cases, it is likely that part of the labor force is covered, but a missing value must be entered because I only use an estimation strategy if at least some information is available. The values of “-99” in the data therefore are true missing data, where there should be a value, but I cannot find enough information to measure it. This contrasts with blank cells in the .csv file (“NA” after importing into an R dataframe) which are missing and there is nothing necessarily expected there. I provide information on imputation and interpolation strategies to apply to the dataset after downloading it in the section ‘Handling missing data and assumptions’.

Figure 1 provides maps of coverage in four important phases of welfare state development – just after the Industrial Revolution, just after World War II, in the 1970s after colonial occupations mostly had ended, and in 2020 (the most recently available measures).

## Figure 1. Four phases of global welfare state development

# Handling missing data and assumptions

The GWIP\_long file does not impute data for years without documented laws or coverage estimates. Only years with at least one source are recorded. Replacement rates and durations are assumed constant from the adoption of a law until a subsequent legislative change. Coverage rates are expected to expand over time but are only updated when sources indicate an increase. For early periods, especially under employer‑liability schemes, coverage estimates are sparse; coverage\_notes and NOTES fields document whether values are inferred (e.g., assumed full coverage within specified sectors) or derived from secondary literature. Users should interpret coverage values as approximate and consult the accompanying codebook for guidance on country‑specific assumptions.

**Quality control**

Data extraction was performed by multiple coders and cross‑checked across sources. Disagreements were resolved through discussion and, where necessary, by consulting additional legal texts. The project maintained a version‑controlled repository for scripts that clean and format the raw entries into the published files. Prior to release, random entries were re‑verified against primary sources. Validation checks ensured consistency of program type classification and chronological plausibility (e.g., the first insurance scheme cannot predate the first law). A list of unresolved ambiguities is documented in the codebook and flagged within the NOTES fields.

**Data Records**

All data files are stored in the Harvard Dataverse repository under persistent identifier DOI: **10.7910/DVN/IVKYIE** and are released under a CC0 Public Domain dedication. GWIP\_v3 (gwip\_v3.tab) provides cross‑sectional data for 189 countries in 2020 with 34 variables, while GWIP\_long (gwip\_long\_v1.0.csv) contains 1,467 country‑year observations for 188 countries and 17 variables. Table 1 summarises the variables in GWIP\_v3. Table 2 summarises the variables in GWIP\_long v1.0. A detailed codebook with theoretical definitions and coding justifications is provided in the file *Breznau and Lanver Global Work‑Injury Policy Database (GWIP) Project Overview and Codebook v2.pdf*[dataverse.harvard.edu](https://dataverse.harvard.edu/dataset.xhtml#:~:text=gwip_long_v1).

**Table 1 – Variables in the GWIP\_v3 cross‑sectional dataset**

| **Variable** | **Description** |
| --- | --- |
| country\_name | Country name. |
| cow\_code | Correlates of War country code. |
| independence | Year of independence or internationally recognised statehood. |
| labor\_workinjury\_firstlaw | Year of the first statutory work‑injury law. |
| labor\_workinjury\_firstlaw\_source | Sources for the year of first law. |
| labor\_workinjury\_firstins | Year of the first insurance scheme (risk pooling or social insurance). |
| labor\_workinjury\_firstins\_source | Sources for the year of first insurance scheme. |
| labor\_workinjury\_first\_fund | Year of the first dedicated accident fund. |
| labor\_workinjury\_first\_fund\_source | Sources for the first fund. |
| labor\_workinjury\_first\_socins | Year when work‑injury benefits were incorporated into a general social insurance programme. |
| labor\_workinjury\_first\_socins\_source | Sources for first social insurance. |
| labor\_workinjury\_firstnat | Year of the first nationwide legislation (applying to entire territory). |
| labor\_workinjury\_firstnat\_carriedover | Indicates whether the national law was carried over from a colonial or regional scheme. |
| labor\_workinjury\_firstlaw\_programtype | Programme type of the first law (employer liability, risk pooling or social insurance). |
| labor\_workinjury\_firstlaw\_programtype\_source | Sources for programme type classification. |
| labor\_workinjury\_firstlaw\_sspw | Year of first law according to the SSPW dataset. |
| labor\_workinjury\_firstlaw\_whitecollar\_fullcoverage | Whether the first law fully covered white‑collar workers. |
| labor\_workinjury\_firstlaw\_whitecollar\_source | Sources for white‑collar coverage. |
| labor\_workinjury\_firstlaw\_bluecollar | Year of first law covering blue‑collar workers. |
| labor\_workinjury\_firstlaw\_bluecollar\_coverage | Level of coverage of blue‑collar workers in the first law (e.g., partial, full). |
| labor\_workinjury\_firstlaw\_bluecollar\_fullcoverage | Whether the first law provided full coverage to blue‑collar workers. |
| labor\_workinjury\_firstlaw\_bluecollar\_source | Sources for blue‑collar coverage. |
| labor\_workinjury\_firstins\_bluecollar\_fullcoverage | Whether the first insurance scheme provided full coverage to blue‑collar workers. |
| labor\_workinjury\_firstins\_bluecollar\_fullcoverage\_source | Sources for blue‑collar coverage under first insurance. |
| labor\_workinjury\_firstlaw\_agriworkers\_fullcoverage | Whether agricultural workers received full coverage under the first law. |
| labor\_workinjury\_firstlaw\_agriworkers\_source | Sources for agricultural worker coverage. |
| labor\_workinjury\_replacement\_rate\_perm\_2020 | Replacement rate (% of wages) for permanent injuries in 2020. |
| workinjury\_duration\_perm\_2020 | Maximum benefit duration (months or unlimited) for permanent injuries in 2020. |
| labor\_workinjury\_replacement\_rate\_temp\_2020 | Replacement rate for temporary injuries in 2020. |
| labor\_workinjury\_duration\_temp\_2020 | Maximum duration (weeks or unlimited) for temporary injuries in 2020. |
| global\_south | Indicator for countries commonly classified as part of the global South (1 = yes, 0 = no). |
| labor\_workinjury\_coverage\_pct\_lf\_2020 | Estimated percentage of the labour force covered by work‑injury insurance in 2020. |
| labor\_workinjury\_coverage\_pct\_lf\_source | Sources for the 2020 coverage estimate. |
| colonial\_history | Historical colonial power or “None” if never colonised. |

**Table 2 – Variables in the GWIP\_long v1.0 longitudinal dataset**

| **Variable** | **Description** |
| --- | --- |
| country | Country name. |
| year | Calendar year of observation. Only years with an applicable law or coverage estimate are recorded. |
| cow\_code | Correlates of War country code. |
| law | Name of the statute or programme in force for the given year. |
| sources | Semi‑colon separated list of primary and secondary sources used for coding that year. |
| type | Programme type in force (employer liability, risk pooling, social insurance or national scheme). |
| NOTES | Additional notes on the law or coding decisions for permanent benefit parameters. |
| labor\_workinjury\_replacement\_rate\_perm | Replacement rate (% of wages) for permanent injuries during this period. |
| labor\_workinjury\_duration\_perm | Maximum benefit duration for permanent injuries (months or “unlimited”). |
| NOTES.1 | Additional notes regarding permanent benefit duration or rate (e.g., lump‑sum payments). |
| labor\_workinjury\_replacement\_rate\_temp | Replacement rate for temporary injuries. |
| labor\_workinjury\_duration\_temp | Maximum duration for temporary injuries (weeks or “unlimited”). |
| NOTES.2 | Additional notes regarding temporary benefits. |
| coverage\_notes | Notes on coverage estimates or assumptions for this year. |
| labor\_workinjury\_coverage\_pct\_lf | Estimated percentage of the labour force covered by work‑injury insurance. |
| labor\_workinjury\_qualifying\_period | Waiting period before eligibility for benefits (e.g., in days). |
| labor\_workinjury\_qualifying\_conditions | Additional conditions required for eligibility (e.g., minimum contributions, specific sectors). |

**Technical Validation**

The accuracy of GWIP\_long depends on the completeness of historical sources and the consistency of coding. To enhance reliability, each entry was independently coded by at least two researchers. Discrepancies were reconciled through discussion and, where necessary, by consulting additional documents. Where multiple sources disagreed on a year or replacement rate, preference was given to primary legislation or official compendia. All sources used are listed in the dataset and codebook, enabling users to verify specific entries. Chronological consistency checks flagged cases where a later program type appeared before an earlier type; these were examined and corrected as needed. A series of automated scripts assessed the range of coverage values, ensuring that replacement rates fell within plausible bounds (0–100 %) and that durations were non‑negative.

Missing data are an inherent limitation. Many early employer‑liability systems did not report replacement rates or coverage, and coverage estimates for the informal sector remain speculative. To minimise imputation bias, GWIP\_long records only years for which there is a documented law or secondary coverage estimate. Replacement rates are assumed constant from the enactment of a law until the next legislative change. Coverage rates likely increase between reforms, but they are only updated when evidence motivates a new estimate. Researchers should therefore interpret the time series as step functions and avoid interpolating trends without consulting the original sources. The coverage\_notes field provides context for each estimate and flags inferred values. A comparison of GWIP\_long against the cross‑sectional 2020 values in GWIP\_v3 shows convergence for countries with stable policies but highlights differences where rapid reforms occurred after 2020, underscoring the need for longitudinal data.

**Usage Notes**

Users can link GWIP\_long with other datasets using the cow\_code and year variables. Because only years with documented laws or estimates are included, researchers performing time‑series analyses should carefully handle missing years and avoid treating the absence of an entry as zero coverage. Coverage estimates (labor\_workinjury\_coverage\_pct\_lf) are approximate; in many countries, actual coverage may lag behind legal entitlements, especially for informal or agricultural workers. The variable names follow a consistent naming convention (labor\_workinjury\_\*) to facilitate merging with other social policy datasets such as the Social Insurance Entitlements Dataset (SIED). The codebook contains additional guidance on harmonising programme types and interpreting qualifying conditions.

When generating visualisations or maps, analysts should account for variation in the years of adoption. For example, a world map shading countries by the decade of first work‑injury law adoption and a line plot showing the cumulative number of adopting countries over time could illustrate policy diffusion. Another figure could display distributions of replacement rates and coverage across programme types, highlighting variation between employer‑liability and social insurance schemes. Users may also wish to produce country‑level trajectories of coverage and generosity. Please avoid including in‑text analyses or conclusions in publications based on this dataset, as *Scientific Data* emphasises that Data Descriptors should focus on documenting datasets rather than presenting scientific results[nature.com](https://www.nature.com/sdata/submission-guidelines#:~:text=For%20Data%20Descriptors%2C%20this%20section,on%20novelty%2C%20impact%2C%20or%20utility).

**Data Availability**

All datasets and supplementary materials for the GWIP project are deposited in the Harvard Dataverse (DOI: 10.7910/DVN/IVKYIE) and are publicly accessible. The repository contains the longitudinal file gwip\_long\_v1.0.csv, cross‑sectional files gwip\_v1.tab, gwip20.tab and gwip\_v3.tab, variable labels files, and the project overview and codebook. The first release of the longitudinal file provides coverage and replacement rates for 188 countries[dataverse.harvard.edu](https://dataverse.harvard.edu/dataset.xhtml#:~:text=gwip_long_v1). Future versions will update the time series as new data become available.

**Code Availability**

All code used to clean, assemble and validate the GWIP\_long dataset is available in a public GitHub repository [link to be provided by authors]. The repository includes scripts for extracting data from source documents, performing quality checks and generating the published CSV and TAB files. No proprietary software was used; all scripts were written in R and Python. Researchers may adapt the code to extend the dataset or to reproduce summary tables and figures.

**Acknowledgements**

The authors acknowledge the German Research Foundation (DFG) Collaborative Research Centre 1342 (“Global Dynamics of Social Policy”) for funding support. We thank the many research assistants who helped gather and code historical laws and the anonymous reviewers whose suggestions improved this dataset. Any remaining errors are the authors’ responsibility.

**Competing Interests**

The authors declare no competing interests.

1. For the first cross-sectional iteration of the GWIP, Felix Lanver and I collected and published the data, for this new longitudinal data I did all of the collection myself. Therefore, the data from the Harvard Dataverse are cited as Breznau and Lanver 5, but this publication is Breznau. [↑](#footnote-ref-1)
2. <https://www.zotero.org/groups/2557302/gwip>. For research purposes, users main join the group and view original legal documents. [↑](#footnote-ref-2)