**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 inception through 2020. The data records the name of the governing law, the type of programme, and the labour force coverage and wage‑replacement rates for both permanent and temporary work-related injury or illness, every time there is a major policy introduced or amended. From the 2000s onward, additional data points are added even in the absence of policy changes because of higher quality source material and with the goal of 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 socioeconomics. I provide R code for users demonstrating how to generate yearly time-series for each country and how to merge with other sources to enable descriptive and statistical analysis and visualizations. A detailed codebook and versioned datasets are available on the Harvard Dataverse and the usage scripts in the Online Repository on GitHub.

# Background

Workers’ compensation is one of the first forms of social welfare policy in human history 1. Its rise coincides with the Industrial Revolution, motivative in part by the horrific injuries and deaths that occurred in the mechanization of extraction and production processes. The idea of protection against risks associated with work has pre-industrialisation origins. Early legal codes in Mesopotamia, Rome and China for example, compensated certain injuries with remunerations, most often for military occupations. The earliest laws put all responsibility in the hands of employers, known as *employer liability*, but this proved to provide little if anything for the workers who were injured in practice. In the nineteenth century Prussian Chancellor Otto von Bismarck introduced a work-injury insurance scheme, creating the first modern *social insurance* programme for industrial injuries. Within a few decades, similar legislation was introduced for other social risks and the ideal of social insurance 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 and prior to the year 2000 was scarce or non-existent until now 4. The *Global Work‑Injury Policy Database Longitudinal* (first release: *gwip\_long* *v1.0*) is a first attempt to fill these scientific data gaps. An earlier cross‑sectional version now in its third release (*gwip v3.1*) harmonised information on the first adoption of work‑injury laws, program types, coverage of various occupational groups and generosity of benefits globally as of 2020. The *gwip v3.1* data are available on the Harvard Dataverse and contain 34 variables for 189 countries (or territories in a few cases).

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. There is a data point entered for each country-year when a major policy change or amendment took place since the introduction of work-injury a policy in each of 188 countries. It includes 1,467 country‑year observations and includes further variables on qualifying periods and conditions, and notes on sources and assumptions made to ensure the highest standard of comparability across countries. A value only exists in the data when primary or secondary sources were found to confirm that a law was introduced or a policy changed. Largely based on the availability of Social Security Programs Throughout the World (SSPTW) data, I coded observational data points every 4 years during the 2000s for most countries.

Data points are missing in between country-years without a known policy change prior to 2000. I do this intentionally to enable the user to make decisions about how best to fill in these data points depending on their analytical goals. As a comparative welfare state scholar, I assume that between observed legislative changes, coverage and replacement rates are relatively stable. Coverage does increase over time based on labour force changes. For example, laws sometimes cover only blue-collar, industrial wage earners. As the percentage of a country’s labour force that are such workers increases due to industrialisation, the coverage rate then increases without a legal change. To generate Figure 1 and XXX in this Data Descriptor, I interpolated coverage values based on basic time and economic growth variable imputations. The workflow to generate the underlying data is provided in the Online Repository. 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) and Comparative Welfare Entitlements Dataset (CWED2)4.

# Conceptual scope

In the GWIP framework a *work‑injury policy* is defined 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. The effectiveness of these work-injury policies in mediating risk and improving the welfare of workers increases substantially across these three types. Work‑injury programmes often laid the groundwork for comprehensive national social security systems2,5.

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 became disabled or if they were born that way. Work-injury is specifically work-related, meaning that some aspect of performing the work caused the injury – that the injury would not have occurred had the person not been working at that specific job. 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, etc. with great variation by country and time period.

The GWIP, both Longitudinal and Cross-Sectional provides *de jure* data. Meaning what is specified by law in the policies. It is entirely possible that a law mandates coverage of all blue-collar workers, but that none are informed of their right, no organizations are formed to oversee that the policy is implemented, and other factors cause a *de facto* coverage of 0%, despite a de jure coverage that is very high.

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, although several smaller countries are included.

# 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. The International Labour Organization (ILO) catalogues many laws in its “Labor Series” and provides them as English translations. Secondary sources included scientific books and articles, the SSPTW series produced by the U.S. Social Security Administration, the ILO’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. For countries that exist today, the COW are analogous alternatives to other well-known coding schemes such as the International Organization for Standardization (ISO) alpha and numeric schemes. In the Online Repository scripts using the R package *countrycode*7 provide mostly easy transfer between such schemes. The COW coding was used because the collection of these data took place as part of a larger project[[3]](#footnote-3) that was interested in the historical divisions of countries that no longer exist today, for example Czechoslovakia, Yugoslavia, Vietnam prior to the Cold War, and etc.

# 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 original and secondary sources. It also documents if the first law was carried over from colonial times or a new enactment. Program types were categorised based on financing and administration: employer liability, provident fund, social insurance or some mixture of these including compulsory insurance on the part of the employee and/or employer. 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 by occupational group.

For the cross-sectional *gwip v3.1*, I extracted the coverage, replacement rates and benefit durations for permanent (1-year) and temporary (6-month) disabilities for the year 2020, or latest available. These are the bridge to the longitudinal data *gwip\_long v1.0* which then extended the data to all years were a source was identified plus every 4 years in the 2000s where possible. There are two slight variations in coding. To get a cross-sectional estimation I relied on a massive data collection effort undertaken by the ILO covering 180 countries for 20208. I was then able to extend these data to 189 countries through imputation (*gwip v3.1*: *labor\_workinjury\_coverage\_pct\_lf\_2020*). However, the data are not historical, and I cannot reproduce exactly the ILO methods as they arrive much closer to de facto coverage (except for undocumented labour of course). Therefore, for the longitudinal data I relied on legal coverage and labour force data, often also take from the ILO to enable a standardized method for over-time comparison. In many cases the results are similar if not identical and in the longitudinal data the exact year is given rather than rounding all years up to 2020, meaning that the latest observation for many countries is 2018 or 2019 (*gwip\_long v1.0*: *labor\_workinjury\_coverage\_pct\_lf*). I compare these two versions in the section ‘Missing data, assumptions and quality control’.

Replacement rates expressing the ratio of wage compensated (*labor\_workinjury\_replacement\_rate\_temp* and *labor\_workinjury\_replacement\_rate\_perm*) were extracted from written law or secondary sources. Table 1 provides some example paragraphs from the original Austrian General Social Insurance Act of September 9th, 19559.

|  |  |  |  |
| --- | --- | --- | --- |
| Paragraph in original act | Relevance | | |
| **4**. *Full insurance*. (1) The following persons shall be insured under the sickness, accident and pension insurance schemes (full insurance) pursuant to this federal Act, unless their occupation is excluded from full insurance by section 5 or 6 or covered only by partial insurance under section 7:  1. workers in the service of an employer or employers;  2. persons bound by a contract of apprenticeship (apprentices);  3. persons intending to enter a profession which requires full university studies or the equivalent who, after completing their studies, are given employment with a view to acquiring the necessary practical training, where they are riot given this training under a contract of employment or apprenticeship (this does not include volunteers);  4. student nurses and student midwives under training in a midwifery training school situated within the federal territory ;  5. persons treated as employed persons within the meaning of subsection (3);  6. homeworkers and persons treated as such (for labour law matters) by virtue of existing legislation respecting home work.  (2) For the purpose of this federal Act, " employed person" means any person employed in a relationship involving economic and personal dependence and receiving remuneration in return; this shall also cover persons in occupations that are characterised more by economic and personal dependence than by independence in the exercise of the occupation. (ILO 1955: Austria 3, pp. 1-2 | | | This defines who qualifies as a “worker” and is therefore covered by this act. This is a relatively standard definition that essentially covers all formal workers. But there are further paragraphs that need to be checked in each case, in case of exemptions. |
| **203**. *Entitlement to accident pension*. The injured person shall be entitled to an accident pension if as a result of an industrial accident or occupational disease his earning capacity is reduced by at least 20 per cent, for more than three months following the occurrence of the event giving rise to benefit; the accident pension shall be payable for such time as the reduction in earning capacity remains at 20 per cent, or more. (ILO 1955: Austria 3, p.54) | I code temporary or permanent inability to work, as at least 80% incapacitated. “Disability” or “incapacity” has varying legal definitions by country, therefore I rely on what the law states. Here 80% is over 20% so this qualifies. | | |
| **204**. *Commencement of accident pension*. (1) If the injured person is entitled under sickness insurance to pecuniary sick benefit on account of incapacity for work resulting from an industrial accident or occupational disease, payment of the accident pension may be claimed as from the day following the day on which payment of the pecuniary sick benefit ceases, or at the latest from the commencement of the 27th week following the occurrence of the event insured against. | | Crucial for data extraction. Sickness provisions cover worker until week 26. | |
| **205**. *Calculation of accident pension*.  (1) The accident pension shall be calculated in accordance with the degree of loss of earning capacity as a result of the industrial accident or occupational disease.  (2) The yearly rate of the pension shall be as follows:  1. 66% per cent, of the basis of assessment (full pension) for such time as the injured person, as a result of the industrial accident or occupational disease, is totally incapable of earning his living;  2. the percentage of the full pension corresponding to the degree of loss of earning capacity (part pension) for such time as an injured person as a result of the industrial accident or occupational disease is partially incapable of earning his living.  (3) As long as the injured person remains unemployed as a result of the industrial accident or occupational disease, the part pension may be increased to the same level as the full pension. (ILO 1955: Austria 3, p. 55) | | Sets the wage replacement rate of 66%. Here I rely on “totally incapable” as referring to someone who is at least 80% disabled (as determined by local law) or otherwise cannot work doing their current job. | |

## Table 1. Original Source Excerpts: Austrian General Social Insurance Act, 1955.

Source ILO Legislative Series. Selected paragraphs from original law, translated by the ILO.

A crucial secondary source of data is the SSPTW. This is useful or checking the data extraction process which in this case is done via qualitative hand coding of the PDFs or online viewable public domain materials. It is also useful for cases where we cannot find an original law. The SSPTW were published first in 1949 by the U.S. Department of Health and Human Services and were updated every three to eight years until 1999 after which the data behind them moved to a data repository as part of the U.S. Social Security Administration and eventually appeared online, updated every four years until 2019. Thereafter the International Social Security Association took over, and provides profiles for 190 countries, and will be the next source for future versions of the *gwip\_long*.

Looking at the SSPTW 1958 book10, which is available online via HaithiTrust, reveals how the data can be extracted. It provides the most up to date information about the benefits, which according to the methods, were extracted via direct communication with social security experts in the government in Austria. Here it lists under “work injury” that the year of “current law” as 1955, and the “Temporary disability benefit” as “Ordinary sickness benefit as above paid for first 26 weeks” (p. 163) and “Permanent disability pension” for “work injury” as 66 2/3 % of average earnings during last year, if totally disabled” (p. 163). There are more provisions relating to child supplement and medical benefits – all things that I found in the original law. These are part of the benefit, but the vary greatly by country and have no standardized form, making them impossible to extract in a standardized way with limited resources. Future researchers could greatly improve the GWIP data by adding variables related to all the additional benefits.

Some countries, in particular early in their histories and especially in the Global South, specify a lump sum for a work-injury. This lump sum had to be divided into the average blue-collar worker pay of that time period in order to covert it into a wage replacement after six months or a year (the coding rules I used for temporary and permanent disability). The coding of laws is further compounded by other laws that alter the definition of worker or create exemptions to the law for certain industries, areas or occupational characteristics. Therefore to get the most accurate extraction, I also searched for these laws. They came in two primary forms. The first relates to colonialism, where colonial powers who enacted laws set up different definitions for natives and Europeans, so that natives were either not considered “workers” by law, or had their own social security systems that differed from Europeans. In these cases calculations were made based on the percentage of European workers in a given country at a time. The other case are Export Processing Zones (EPZs)

# Primary variables of interest

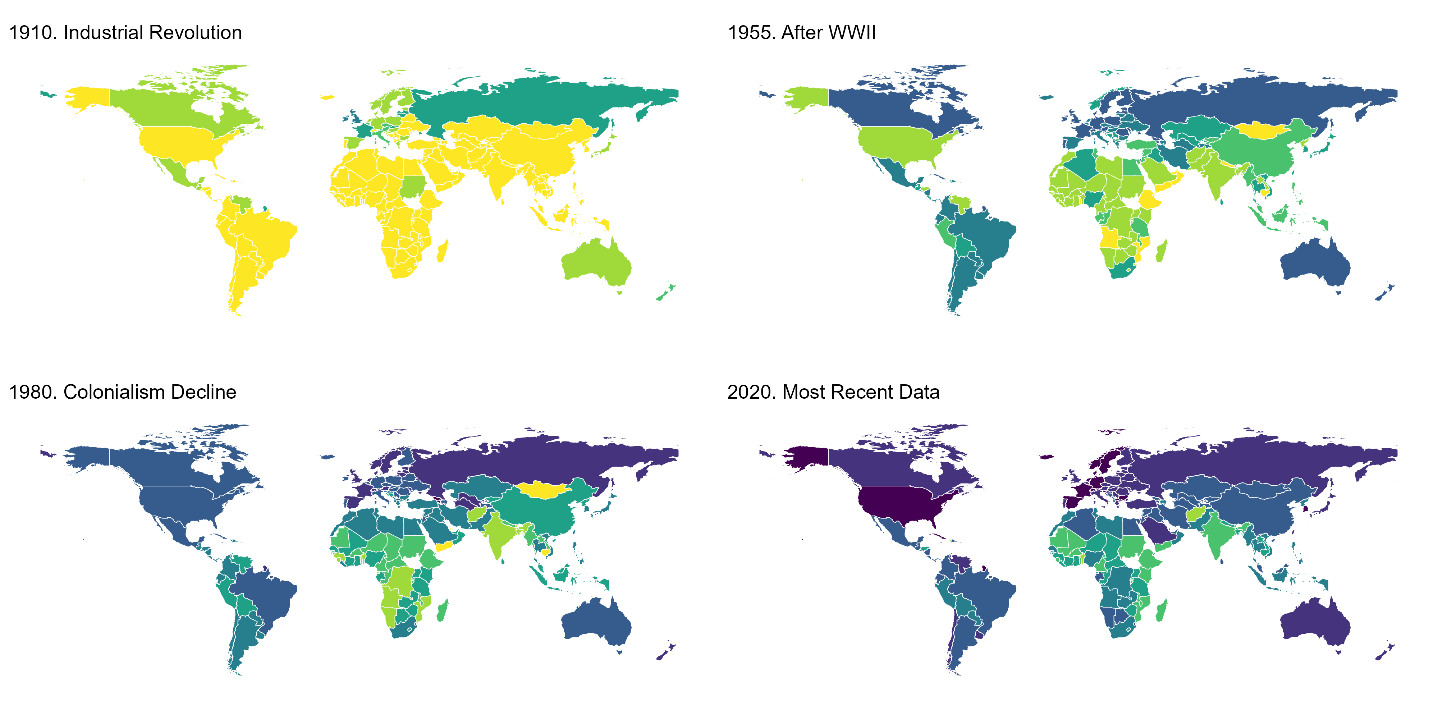
Detailed definitions and further information about all variables in both the Cross-Sectional and Longitudinal GWIP are available in the Codebook 11. Here I will go into detail only regarding two central variables: coverage (*labor\_workinjury\_coverage\_pct\_lf*) and temporary and permanent injury wage replacement rates (*labor\_workinjury\_replacement\_rate\_temp* and *labor\_workinjury\_replacement\_rate\_perm*). 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 12–15.

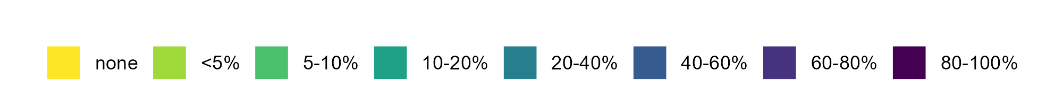
Coverage as a percentage of the labor force is measured as those who are legally covered out of the formal labor force. This excludes informal and shadow economy work, and subsistence activities like farming. Legal coverage comes at first from the legal text. If a law states that all “workers” are covered, 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 20204,16. 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. Coverage rate of work-injury protections as a percent of the labor force across four phases of global welfare state development

Data are from the gwip\_long v1.0 with interpolation between policy observations, see [Workflow Repository](https://github.com/nbreznau/gwip_long_user).

## Figure 2. Replacement rates

# Missing data, assumptions and quality control

Distinguishing missings: A -99 means there is a law and therefore should be a value, but we could not find it. A “0” means that there is a value, but only miners are covered or that there was only compensation for death, and this is not enough to count as a full law or real injury compensation. A “none” means that there was no law covering this but there was some other reason to have a data point here.

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.

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.

## Figure 3. Comparing GWIP and ILO coverage rate as a percentage of the labour force

**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 *Global Work‑Injury Policy Database (GWIP) Project Overview and Codebook v2* 11published together with my co-researcher Felix Lanver whose participation included the development of the gwip v1.0 Cross-Sectional data and supporting materials, which therefore keeps him as a co-author on the newer version of the codebook which I otherwise alone updated.

**Table 2 – 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 SSPTW 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 3 – 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 6, 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)
3. The data were collected as part of the University of Bremen’s Comparative Research Centre 1342 “The Global Dynamics of Social Policy”. <https://www.socialpolicydynamics.de/en/> [↑](#footnote-ref-3)