

CODEBOOK AND MANUAL FOR GATHERING AND ENTERING DATA IN THE DATABASE OF THE GLOBAL COLLABORATORY ON THE HISTORY OF LABOUR RELATIONS 1500-2000

International Institute of Social History; <http://www.historyoflabourrelations.org/>

Version 12 (24/05/2011): Karin Hofmeester, Mieke Mertens, Kees Mandemakers, Christine Moll Murata, Lotte van der Vleuten

Version 13 (14/1/2014): Rombert Stapel

Version 14 (27/10/2015): Rombert Stapel

CONTENTS

Contents .....	1
Introduction .....	1
Instructions for using the blank databases (Access and Excel) .....	3
Table 1 – Population.....	4
Generally accepted combinations of labour relations.....	7
Screenshots for entering labour relations .....	8
Table 2 – Sources.....	10
Table 3 – Typology IV .....	10

INTRODUCTION

This database is part of the Global Collaboratory on the History of Labour Relations 1500-2000. This project aims to make an inventory of all types of labour relations worldwide, varying from slavery, indentured labour, sharecropping to free wage labour and self-employment in all its facets and combinations and to make a reconstruction of the developments of these types from 1500 up until today. To get an insight in the worldwide development from 1500 to now, we gather data on the occurrence of all types of labour relations in all parts of the world during five cross-sections in time i.e.: 1500, 1650, 1800, 1900 and 2000 (and 1950 for Africa). To do so we asked regional experts from different disciplines to co-operate with us, gathering and sharing their data and its critical annotations in a collaboratory and during meetings. The creation of this research infrastructure should lead us to the discovery of lacunae in our knowledge and analysis which can be solved through further research.

The main, short term goal of the database is to create an instrument that will show us which types of labour relations prevailed in the world, in the various cross-sections. To do so, we at least need the total population of a 'nation'. To know/estimate/guestimate which part of this population, e.g. how many individuals, worked in each of the labour

## Codebook and Manual for Gathering and Entering Data

relation types, we deduct those individuals that we consider non-working.<sup>1</sup> Of the remaining individuals, that by default are considered working, we would like to know their occupation. Gaining knowledge about the occupational structure is one of the intermediate stages to gaining insight in the occurrence of labour relation types. By showing how we attached types of labour relations to occupations, we show how our inventory has been made. This also serves the long term goal of the database.

In the text below, we will suggest the preferable geographical level of the data, as well as the level of standardization of occupational titles. If data is aggregated in your sources, and these aggregations are trustworthy, please use the highest aggregation level if this is workable. In the text below, we will follow the order of the data as they should be entered in the database, and therefore the subdivision of the codebook of the database.

---

<sup>1</sup> The non-working population is divided into three labour relations: 1) Those who cannot work or cannot be expected to work (because they are either too young, too old, disabled or studying); 2) Those who do not have to work (because they are affluent); 3) Those who wish to work but have no job (unemployed).

### INSTRUCTIONS FOR USING THE BLANK DATABASES (ACCESS AND EXCEL)

For each combination of country and cross-section in time a separate database should be produced. In order to maximize the uniformity and interoperability of the databases and in order to minimize errors a set of precautionary measures have been taken.

First, for each record, every field has to display content. The only exception is the field 'Id\_labrel' which is filled in by the database administrator. Not all fields will be applicable all the time. In practice, this means that for many fields a code for "not applicable" must be used. Depending on the field this code can be "NA", "-1", "-99.999.998" or "0%". Some fields will not have the option "not applicable" or an equivalent thereof. These are: 'Id\_pop', 'Year', 'Country', 'Quality\_total', 'Gender', 'Mar\_stat', 'Age\_start', 'Age\_end', 'Type\_activity' and 'Source'. Details can be found below.

To facilitate the user, we have created two specially prepared files in which the data can be entered. Both files will give immediate feedback whether the data is entered correctly. Available for all users are two Excel files (Excel 2003 and Excel 2007 and up. Both are available on our website or GitHub (rlzjdemman/labrel). The Excel files are designed to be as user-friendly as possible, giving real-time feedback on the correctness of their data entry (following our rules for data entry).

The Excel file will not automatically accept all possible combinations of labour relations (e.g. "14018") or HISCO major group codes (e.g. "7/8/9") (see below for more information regarding such combinations). The generally allowed combinations of labour relations have been added, but if one wishes to include certain unique combinations (HISCO or labour relations), one may add these manually in the appropriate column in the Excel worksheet tab "Labour relations used".

In the Excel file the field 'Source' is linked to the 'Id\_source'-number that should be appointed to each source in the appropriate Excel tab.

For those who want to use Access rather than Excel, this opportunity is also available, but it allows the user to make mistakes in the data entry much more easily. Please save the database using the file format ".mdb" rather than the newer format ".accdb", which cannot be opened by older versions of Access.

NB. Note that some HISCO combinations (such as 5/6) are automatically changed into dates by Excel. This problem can be solved by adding an apostrophe at the beginning: '5/6.

**NB. Please be aware that the conditional formatting in Excel is affected by 'cutting and pasting' or deleting of cells. Use only 'copy and paste (as values)' when entering sets of data this way. All cells that should not normally be altered have been secured (which can be turned off if necessary in the settings of Excel).**

TABLE 1 – POPULATION

In this table the unit of research is the total number of people for a particular cross section in time, in a particular area. This population is characterised by several attributes, e.g. gender, age, marital status, economic branch, occupation, nationality, ethnicity, religion, social group. The data are entered as they are found in the sources, the sources are mentioned in Table 2. To the data found in sources an interpretation is added, i.e. labour relation. The typology of this feature is given in Table 3.

Please state the total population number as much as possible on a national level, even if you have to estimate/guestimate. If certain regions are important for your area in your cross section(s), please enter also the total population number on this geographical level. The same goes for the total population of important cities, that remained more or less important over a longer period. When there is local or regional data on labour relations available that is important to include, please also include as much local or regional data as possible on gender, marital status, age distribution, etc.

Field name	Allowed data	Description	Explanation and requirements
<b>Id_labrel</b>		Identification number for IISH database administrator	Leave blank
<b>Id_pop</b>	#	Identification number	Primary key, must be unique. Enter manually
<b>Day</b>	-1	Day	Infrequently used feature, included for backward compatibility
<b>Month</b>	-1	Month	Infrequently used feature, included for backward compatibility
<b>Year</b>	0-2020	Year	Can be estimate or average (specified under 'Remark')
<b>Year_start</b>	-1	Start period	Infrequently used feature, included for backward compatibility
<b>Year_end</b>	-1	End period (max. 25 years)	Infrequently used feature, included for backward compatibility
<b>Locality</b>	Text/NA	Name of locality	
<b>Urbanized</b>	Y/N/NA	Urban population	Y for part of total population that lives in localities with > 5.000 inhabitants; N for part of total population that lives in localities with < 5.000 inhabitants. If your sources give another definition of a city (for example >3.000), please state so in under 'Remark'
<b>Region</b>	Text/NA	Name of region	
<b>Country</b>	Text	Name of country	
<b>Total</b>	#/-99.999.998 <sup>2</sup>		It is mandatory to include a number in 'Total'. If no number is available, an estimate, guestimate or a calculated average of the high and low estimates may be given. Please specify this under 'Remark'. High and low estimates can be entered in the fields 'Minimum' and 'Maximum'.
<b>Minimum</b>	#/-1	Minimum estimation	
<b>Maximum</b>	#/-1	Maximum estimation	
<b>Quality_total</b>	A/C/E	Code indicating the quality of the data	A=Aggregated, C=Calculated, E=Estimated. If estimated by yourself: please state so in remark field and explain how it is done; if estimation is found in another source, indicate the source in the source field
<b>Gender</b>	T/M/F/U	Gender	T=Total, M=Male, F=Female, U=Unknown

<sup>2</sup> This number was chosen so that the project administrator is able to quickly detect when these numbers are wrongfully included when automatically calculating sums.

## Codebook and Manual for Gathering and Entering Data

			Please note that when gender is split up, the total (sum of both genders) should also be given
<b>Mar_stat</b>	T/M/S/W/D/U	Marital status	T=Total, M=Married, S=Single, W=Widowed, D=Divorced, U=Unknown Please note that when marital status is split up, the total (sum of marital statuses) should also be given
<b>Age_start</b>	0-[98]	Start date age group	Lower open end = 0, higher open end = 99.
<b>Age_end</b>	[1]-99	End date age group	Use the age groups your sources are indicating, we can standardize afterwards. Keep in mind though that the age distribution for labour relation 1 is already set ("Cannot work or cannot be expected to work: those who cannot work, because (they are too young ( $\leq 6$ years), too old ( $\geq 75$ years)") See our <i>Definitions of Labour Relations</i> , p.1). If the age people start and stop working differ strongly in your area, please state so in under 'Remark' and in the methodological paper.
<b>Type_activity</b>	P/L (+a/c/d/j)	Code used to distinguish data regarding 'Population' and 'Labour Relations' in the database	P=Population, L=Labour Relation. Can be combined with: a= excluding part of the area; to be defined under 'Remark' c= excluding (part of) (Indian/aboriginal/ Maori) population d= excluding part of population; to be defined under 'Remark' j= excluding (part of the) military personnel/armed forces/persons on compulsory military service
<b>Branch</b>	Text/NA	Branch economic active population as stated in your source	
<b>Branch_HISCO</b>	0-9/NA	HISCO Major Group code that corresponds with 'Branch'	See: <a href="http://historyofwork.iisg.nl/major.php">http://historyofwork.iisg.nl/major.php</a> , major group 0 to 9 <sup>3</sup> Two different HISCO branch codes should be separated with a slash. The HISCO Branch code is only one number long.
<b>Occup</b>	Text/NA	Occupational title as stated in your source	
<b>Occup_HISCO</b>	010-999/NA	HISCO Minor group code that corresponds with 'Occup'	Please enter the 3 digit unit group code, since this makes adding information on class and status easier. If only two digits are available, please add a zero at the end.
<b>Position</b>	Text/NA	Status as indicated in the source	Concerns the individual

<sup>3</sup> Please note that though HISCO clusters some of these major groups in their overview on the website, these should be entered separately in the database. The major groups of HISCO are as follows:

*Major group 0* = professional workers, technical and related workers

*Major group 1* = professional workers, technical and related workers

*Major group 2* = administrative and managerial workers

*Major group 3* = clerical and related workers

*Major group 4* = sales workers

*Major group 5* = service workers

*Major group 6* = agricultural, animal husbandry and forestry workers, fishermen and hunters

*Major group 7* = production and related workers, transport equipment operators and labourers

*Major group 8* = production and related workers, transport equipment operators and labourers

*Major group 9* = production and related workers, transport equipment operators and labourers

## Codebook and Manual for Gathering and Entering Data

<b>Nationality</b>	Text/NA	Nationality	
<b>Ethnicity</b>	Text/NA	Ethnicity	
<b>Race</b>	Text/NA	Race	
<b>Religion</b>	Text/NA	Religion	
<b>Social_group</b>	Text/NA	Social group	i.e. tribe, caste
<b>Source</b>	1-1000	Secondary key to source table	Corresponds to 'Id_Source' in Table 2 (or in the Excel tab 'Sources' when applicable).
<b>Volume</b>	#/Text/NA	Number of volume source	
<b>Page</b>	#/Text/NA	Page number source	
<b>Labour_rel_01</b>	-1/1 ... 18/NA	Type of Labour Relation, interpreted by researchers <sup>4</sup>	<p><b>Indefinable labour relations:</b> one may combine two or more codes <i>only</i> (!) if you do not know exactly into which category your total belongs. For example: you know the number in 'Total' refers to wage labourers, but you do not know if they work for the market or for non-market institution. Enter 14018 in 'Labour_rel_01'. <i>See the list below of generally accepted combinations.</i></p> <p><b>Multiple labour relations:</b> Use 'Labour_rel_02' and 'Labour_rel_03' if you know that the people that make up the number in your 'Total' combine different types of labour relations. Enter the main labour relation in 'Labour_rel_01', and the secondary in 'Labour_rel_02', or even the third in 'Labour_rel_03'.</p> <p><b>Percentages:</b> Estimate the percentages of time spend in a year for the different labour relations. If you know the Total number works as wage labourers for the market for <b>9 out of 12 months</b>, and is self-employed for the other 3 months, enter 14 in 'Labour_rel_01' and 75% in 'Labour_rel_01_%'. Enter 12 in 'Labour_rel_02' and 25% in 'Labour_rel_02_%'.</p> <p>Percentages should add up to 100%. Use "0%" when the accompanying 'Labour_rel_0X' states "NA".</p> <p><b>Leftover categories:</b> If you have a 'leftover' category for which you have no other information on Labour Relations, give it the Labour Relation -1 and name this Labour Relation "Unknown" in 'Occup' and/or 'Branch'. This figure should be very small, and never higher than 1% of the total population.</p>
<b>Labour_rel_01_%</b>	1-100%/0%		
<b>Labour_rel_02</b>	-1/1 ... 18/NA		
<b>Labour_rel_02_%</b>	1-100%/0%		
<b>Labour_rel_03</b>	-1/1 ... 18/NA		
<b>Labour_rel_03_%</b>	1-100%/0%		
<b>Remark</b>	Text/NA	Remarks	

<sup>4</sup> See our *Definitions of Labour Relations*, <http://bit.ly/1doPk5t>.

---

GENERALLY ACCEPTED COMBINATIONS OF LABOUR RELATIONS

Combinations of labour relations (e.g. 14018) are meant as an escape route whenever a lack of sources prevent the use of more specific labour relations. However, not all combinations of labour relations that are theoretically possible are recommendable in practice. Only a few combinations are generally accepted – more often than not this is a question of common sense. Please note, that even the use of these combinations should be avoided as much as possible. The combinations that are allowed are:

**12a012b\*, 12a013\*\*, 12b013\*\*, 12a012b013\*\*, 12a014, 12b014, 12a012b014, 12a018, 12b018, 12a012b018, 12a013014\*\*, 12b013014\*\*, 12a012b013014\*\*, 12a014018, 12b014018, 12a012b014018, 13014\*\*, 14018, 4a04b\*, 4a04b012a012b\*, 4a012a, 4b012b, 105\*\*\***

*\*) If possible, use information on household size to estimate the ratio between 4a/12a and 4b/12b.*

*\*\*\*) If possible, find sources (e.g. business counts) to estimate the ratio between 12a/12b/14 and 13. Normally, this should never fall short of a ratio of 4:1.*

*\*\*\*) In some cases, often females, the categories “cannot work or cannot be expected to work” and “household kin non-producers” are not easily distinguishable in certain sources, such as old fashioned census records. However, this combination should be avoided at all cost.*

NB. The use of combinations, especially if one has compelling reasons to deviate from the aforementioned list, often caused by unique characteristics of the sources, should always be discussed explicitly in the methodological paper.

## Codebook and Manual for Gathering and Entering Data

### SCREENSHOTS FOR ENTERING LABOUR RELATIONS (ACCESS)

The screenshot shows the Microsoft Access interface with the 'Table Tools' ribbon active. The 'Labour\_rel\_01' column is selected in the data table. The table has columns: Page, Labour\_rel\_01, Labour\_rel\_01\_%, Labour\_rel\_02, Labour\_rel\_02\_%, Labour\_rel\_03, and Labour\_rel\_03\_%. The data rows are:

Page	Labour_rel_01	Labour_rel_01_%	Labour_rel_02	Labour_rel_02_%	Labour_rel_03	Labour_rel_03_%
1						
2						
4						
5						
12						
14						
*						

Figure 1. Entering labour relations in 'Labour\_rel\_01'.

The screenshot shows the Microsoft Access interface with the 'Labour\_rel\_01' column selected in the data table. The value '14018' is entered in the cell for the row with Page '12'. The table has columns: Page, Labour\_rel\_01, Labour\_rel\_01\_%, Labour\_rel\_02, Labour\_rel\_02\_%, Labour\_rel\_03, and Labour\_rel\_03\_%. The data rows are:

Page	Labour_rel_01	Labour_rel_01_%	Labour_rel_02	Labour_rel_02_%	Labour_rel_03	Labour_rel_03_%
1						
2						
4						
5						
12	14018					
*						

Figure 2. Entering labour relations when in doubt.

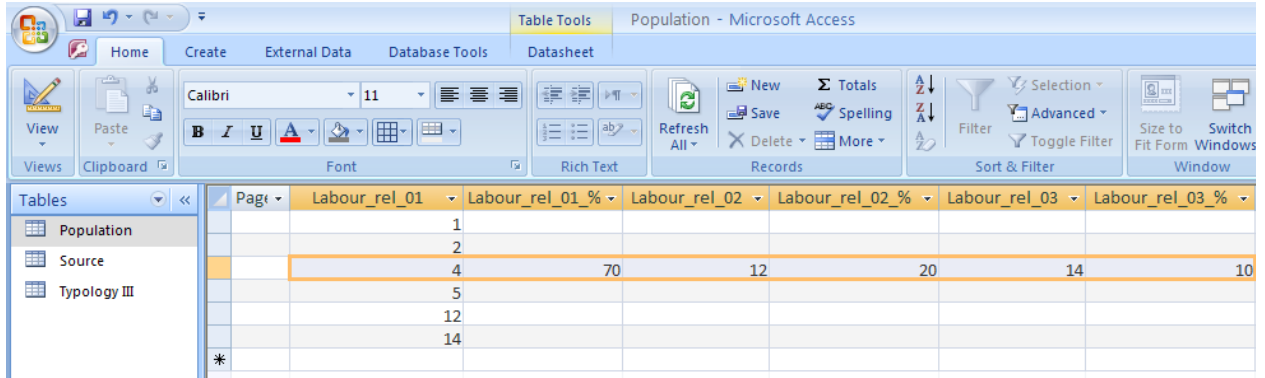
The screenshot shows the Microsoft Access interface with the 'Labour\_rel\_01' column selected in the data table. The values '4', '12', and '14' are entered in the cells for the rows with Page '4', '12', and '14' respectively. The table has columns: Page, Labour\_rel\_01, Labour\_rel\_01\_%, Labour\_rel\_02, Labour\_rel\_02\_%, Labour\_rel\_03, and Labour\_rel\_03\_%. The data rows are:

Page	Labour_rel_01	Labour_rel_01_%	Labour_rel_02	Labour_rel_02_%	Labour_rel_03	Labour_rel_03_%
1						
2						
4	4					
5						
12	12					
14	14					
*						

Figure 3. Entering combined labour relations.



## Codebook and Manual for Gathering and Entering Data



The screenshot shows the Microsoft Access interface in Datasheet view for a table named 'Population'. The ribbon includes 'Home', 'Create', 'External Data', 'Database Tools', and 'Datasheet'. The 'Home' ribbon is active, showing options for font, rich text, records, and sort & filter. The table has seven columns: 'Page', 'Labour\_rel\_01', 'Labour\_rel\_01\_%', 'Labour\_rel\_02', 'Labour\_rel\_02\_%', 'Labour\_rel\_03', and 'Labour\_rel\_03\_%'. The data is as follows:

Page	Labour_rel_01	Labour_rel_01_%	Labour_rel_02	Labour_rel_02_%	Labour_rel_03	Labour_rel_03_%
1						
2						
4		70	12	20	14	10
5						
12						
14						
*						

Figure 4. Entering combined labour relations with yearly share of time in percentages.

TABLE 2 – SOURCES

All sources used to build the database are entered in a separate table that corresponds with the “Source” field in the main population table. In this table, all sources used to build table 1 are described and encoded. Codes refer to paper copies of essential parts of the sources.

Field name	Allowed data	Description
<b>Id_Source</b>	1-1000	Identification number for the source
<b>Author</b>	Text	First name and last name of author
<b>Title</b>	Text	Full title
<b>Place/Year/Volume</b>	Text	Place, year and volume of publication
<b>Nr_Copy</b>	#	Code referring to paper copies of the sources
<b>Remark</b>	Text	Remarks

TABLE 3 – TYPOLOGY V

This table gives an overview of all prevailing labour relations, as developed by the research group (version V, October 2015). The several types are determined by the ownership of labour power; labour instruments; output and means of subsistence. **This table is for reference only.**

Field name	Allowed data	Description
<b>Id_type</b>	-1/1 ... 18	Identification number of labour relation type
<b>Definition</b>	Text	Definition
<b>Working for</b>	Text	Institution workers produce for
<b>Type of Labour</b>	Text	Main type of labour relation