The CORESIDENCE LA Database:
Living Arrangements Around the World, 1970-2021

Juan Galeano[1], Albert Esteve[1]

1: Universitat Autònoma de Barcelona / Centre d'Estudis Demogràfics

Corresponding autor: Juan Galeano (jgaleano@ced.uab.es)

Abstract

This paper introduces the CORESIDENCE Living Arrangements Data Base (CoLADB), a novel global database designed to address the significant gap in statistical information available for analysing the composition, distribution, and evolution of living arrangements over time. Utilizing comprehensive microdata from IPUMS International and the European Labour Force Survey (EU-LFS), CoLADB encompasses over 782 million individual records from 107 countries, spanning the period from 1970 to 2020. The database employs an innovative algorithm to reconstruct kinship relationships within households, providing a robust and scalable methodology for examining living arrangements. Additionally, CoLADB integrates kinship estimates from the Projections of Human Kinship project, offering deeper insights into the demographic landscapes of different regions. The development of CoLADB involved a collaborative effort with the Barcelona Supercomputing Centre, leveraging the computational power of MareNostrum V to process the extensive data. This resource is expected to be instrumental for researchers and policymakers, enabling evidence-based decision-making in areas such as housing, social services, and healthcare. The open-source R code used in this project is publicly available, promoting transparency and facilitating the development of new living arrangement typologies for diverse research goals.

Background & Summary

Our lives are comprised of moments shared with others or spent in solitude. Regardless of the circumstances, many of these moments occur within the confines of our household. The people with whom we live, as well as their absence, significantly influence who we are and the opportunities we have in life[1]. At every stage of the life cycle, our identities—whether as parents, children, siblings, partners, friends, or roommates—along with our personalities and well-being, are deeply influenced by our living arrangements[2,3]. These arrangements not only impact us individually but also mirror and reinforce societal norms that structure how we organize ourselves collectively [4,5]. From extended family forms find in parts of Asia, African and Latin-American to nuclear families and unipersonal households in Europe and North America, living arrangements reflects the vast heterogeneity of human experience.

Living arrangements, in turn, are affected at the macro level by shifts in demographic and economic trends and changes in social norms but also influenced, at the micro level, by social policies aiming to tackle these changes[6,7,8]. Understanding these arrangements provides valuable insights into a key component of the social fabric: our social support systems, which encompass the provision of care and well-being and represent our fundamental network of socialization. This knowledge not only enhances our comprehension of societal dynamics but also informs policy-design in critical areas such as housing, education, social services, and healthcare. Despite the central role households play in social reproduction and change, global statistical data on the composition and evolution of living arrangements remains limited. The CORESIDENCE Living Arrangements Data Base (CoLADB) seeks to bridge this gap by offering a worldwide database dedicated to this topic. By doing so, it also introduces a robust and scalable methodology for reconstructing living arrangements within households based on population data repositories and surveys.

CoLADB has been constructed utilizing comprehensive microdata from two major repositories accessible to researchers: IPUMS International and the European Labour Force Survey (LFS). Each country-year sample included allows to group individuals into households and facilitates the analysis of kinship relationships, or their absence, among household members. Furthermore, the microdata provides essential sociodemographic details about household members, such as age, sex, educational attainment, and marital status. The aggregated database presented in this article is derived from complete samples available within IPUMS-I and LFS, encompassing over 782 million individual records and spanning 107 countries from 1970 to 2020.

With whom do we live is constrained, to a greater extend, by the availability of kin. Over the last decades, two major demographic trends have shaped the global landscape: a continuous and widespread decline in fertility rates[9,10,11] and an increase in life expectancy[12]. The interaction between these trends is expected to have profound implications for the structure of family networks, not only due to the decreasing number of available kin but also because of the aging of these kin networks. These projected changes in kinship structures underscore the growing need for robust social support systems. As family networks become smaller and older, individuals will have fewer relatives to depend on for informal care, leading to increased demand for formal social support systems, particularly in regions where welfare systems are underdeveloped. For this reason, the CoLADB includes, as external data, the estimates provided by the Projections of Human Kinship for All Countries produced by Diego Albúrez-Gutierrez and colleagues[13].

Additionally, the open-source R code used in this project is publicly available, enabling users to scrutinize how the microdata was processed and how the indicator on living arrangements was developed. This transparency promotes replicability and empowers users to construct new living arrangement typologies suitable for their particular research goals. CoLADB has been developed within the project "Intergenerational Coresidence in Global Perspective: Dimensions of Change (CORESIDENCE)", funded by the European Research Council. It is designed to complement and expand the research possibilities in the field of family and household studies present in our previous database on national and subnational data on the size and composition of households around the world[14].

CoLADB opens up vast research possibilities, allowing users to analyse the distribution of living arrangements by factors such as age, sex, educational attainment, and marital status across different countries and years. It enables researchers to address fundamental questions like: Who do we live with throughout various stages of life? How much time do we spend in different family setups? And what are the key factors driving the transitions in household compositions over a lifetime? By offering detailed insights into the dynamics of living arrangements, CoLADB helps to uncover the patterns and determinants that shape household structures and their evolution over time, fostering a deeper understanding of social and demographic trends globally.

1. Methods

1.1 Overview

Figure 1 provides a schematic overview of the entire process of creating CoLADB, starting with data acquisition, and followed by data processing, reconstruction of the living arrangements for all household members, harmonization of the variables of educational attainment and marital status, output datasets, and external validation.

CoLADB draws on two main repositories of global-scale individual microdata: The International Integrated Public Use Microdata Series (IPUMS-I) and the European Union

Labour Force Survey (EU-LFS). The estimates of available kinship come from the Projections of Human Kinship for All Countries.

All data cleaning, processing, harmonization, and aggregation were conducted using R[15]. The complete code for constructing CoLADB is available in the project's GitHub repository (see section **Code Availability**). The vast scale of this project—encompassing over 787 million individual records and more than 3TB of data—along with the substantial computational power required to implement the algorithm for reconstructing living arrangements within each household, exceeds the capabilities of a standard computer for performing the necessary tasks. To address these computational challenges, we partnered with the Barcelona Supercomputing Centre (BSC), home to MareNostrum V—the most powerful computer in Europe and the eighth most powerful in the world. This collaboration, which bridges Social and Computational Sciences by integrating computer engineering, parallel computing, and demographic analysis, has enabled us to process the extensive data underlying CoLADB.

The output data of CoLADB consists of four distinct datasets: Single Ages, Age Groups for both IPUMS and LFS data, and a Harmonized dataset.
The Single Ages dataset includes 301 country-year samples from 91 countries worldwide (Fig. 2), constructed using over 760 million individual records from the full samples available in IPUMS-I. This dataset aggregates original microdata by single age, educational attainment, and marital status—using the categories defined by IPUMS-I—and by a set of types of living arrangements (see section 1.3 **Living Arrangements**). It also includes estimates of available kinships.
The five-year age groups datasets are derived from 319 country-year samples from IPUMS-I and 86 from LFS, covering 787 million individual records across 107 countries worldwide. These datasets maintain the original categories for educational attainment and marital status as defined by their respective sources, with additional estimates of available kin.
Finally, the Harmonized dataset provides information aggregated into five-year age groups (for single ages IPUMS-I samples) and by types of living arrangements, marital status and educational attainment. In this dataset, categories of educational attainment and marital status have been harmonized (see section 1.4 **Harmonization process**). Like the other datasets, it also includes estimates of available kin.

To ensure the accuracy and reliability of the CoLADB, we validated our database by comparing the results of a selected set of living arrangements with their corresponding household types constructed by IPUMS-I and included in each of the samples (see section 3 on **Technical Data Validation**).

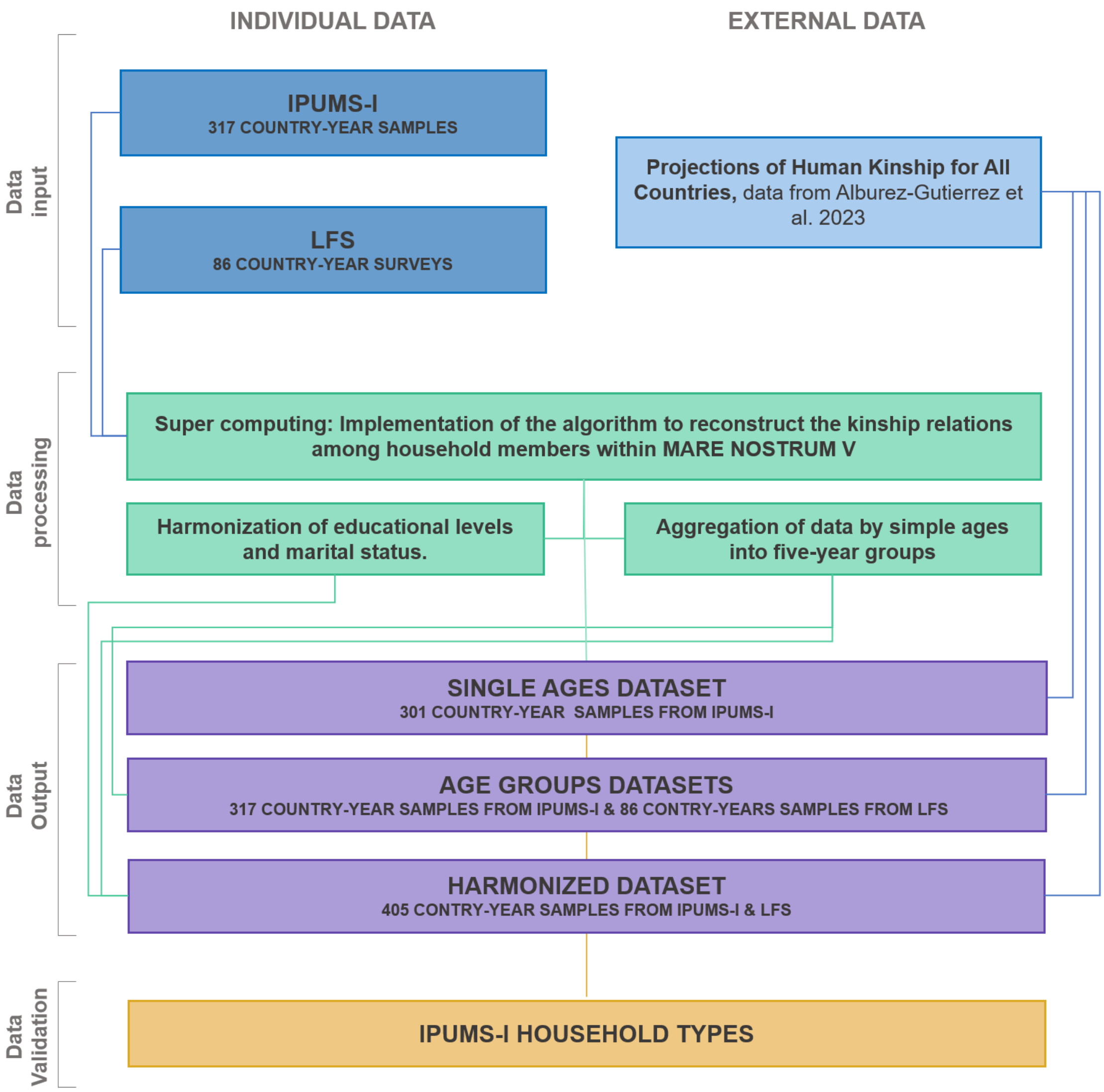

Fig.1: Flowchart representing the different stages to build the CoLADB

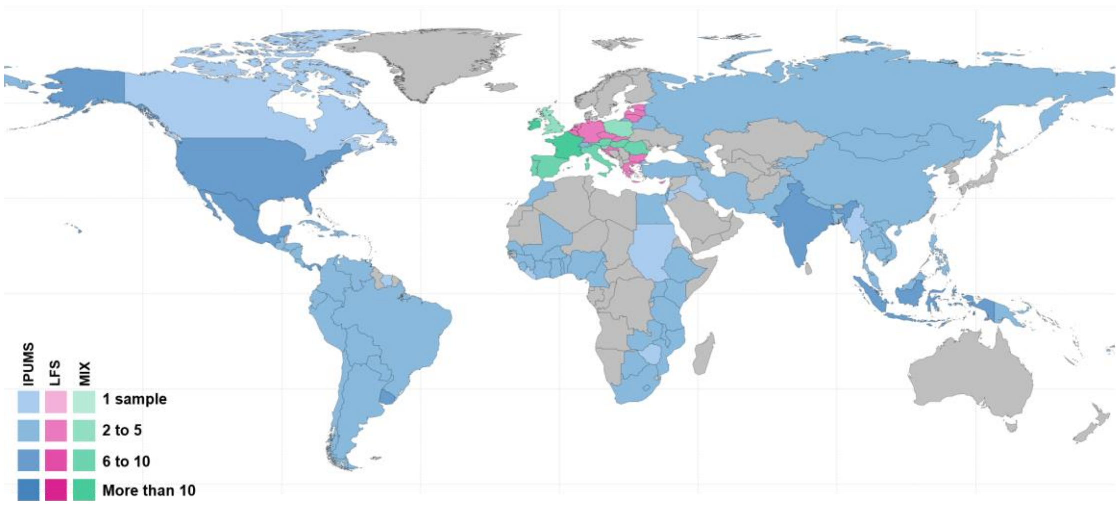

Fig. 2: Country coverage by number of samples available of the CoLADB

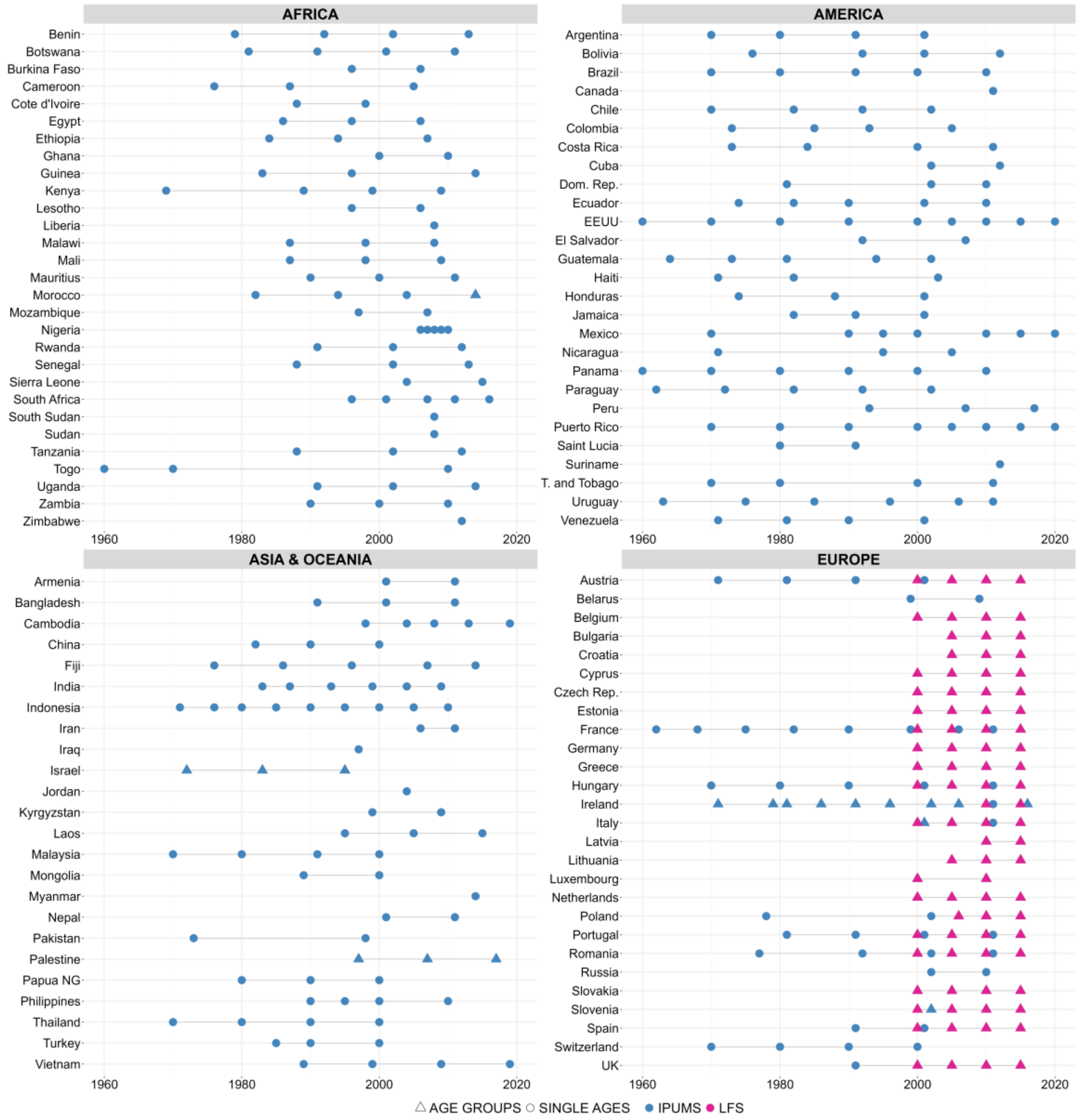

Fig. 2: Availability of samples by country, year, ages and source in the CoLADB

1.2 Data Sources

The CoLADB is a comprehensive source of information on living arrangements at the national level. The database draws on two major repositories of individual microdata on a global scale. Additionally, we employed external data to provide a set of indicators on the availability of kin.

The primary source of individual microdata for CoLADB is the International Integrated Public Use Microdata Series (IPUMS-I)[16], which consists of 316 census samples from 94 countries (https://international.ipums.org). The IPUMS International project is a global initiative dedicated to the collection, preservation, harmonization, and distribution of census microdata from across the globe. For the construction of CoLADB, we utilized the complete samples available in IPUMS-I, encompassing more than 782 million individual observations. This extensive dataset provides a robust foundation for analysing living arrangements trends at a global scale.

In order to complements the scarce information available on European countries from IPUMS, we include, as secondary source of individual microdata, 86 samples of the European Labour Force Survey (EU-LFS)[17]. (https://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey). The EU-LFS is a large household sample survey on the labour force participation of the 15-year and older population, also collecting information on all members of the household surveyed, as well as the kinship relations among them. Crucially for our porpoises, samples from year 2000 onwards include the necessary pointer variable for the reconstructions of living arrangements. As LFS collects data on a quarterly basis, samples included in the CoLADB correspond to the yearly samples to ensure consistency with the specific time frame for which the data was downloaded.

The CoLADB has been designed with a forward-looking perspective, poised to accommodate the ongoing growth of its constituent data repositories. As the aforementioned data sources continue to release new samples, the CoLADB is primed to seamlessly integrate these additions, ensuring its comprehensiveness over time.

1. 3 External data

With whom do we live is constrained, to a greater extend, by the availability of kin. Taking these into consideration we decided that CoLADB would benefit of providing the estimates of available kin produced by Alburez and colleagues[13], and these are included in all datasets of this new database. The estimates are provided in five-year ages groups and include information on the availability of aunts/uncles, cousins, children, grandchildren, great-grandchildren, grandparents, great-grandparents, parents, sibling and niblings (nieces and nephews). By including the estimates of available kin, the CoLADB empowers researchers and policymakers to explore the demographic landscapes of different countries and time periods in relation to changes in living arrangements. These indicators facilitate a deeper understanding of population dynamics, thereby supporting evidence-based decision-making and policy formulation.

1.4 Living arrangements

One of the contributions to the creation of CoLADB has been the development of an algorithm that reconstructs kinship relationships between household members by assigning a living arrangement value to each individual. This algorithm operates as a numerical encoding system that detects kinship based on nine potential relationships to a reference individual, or "ego": father, mother, child, partner, sibling, grandparent, grandchild, other relative, or non-relative. These relationships are encoded using powers of two for calculation. The algorithm relies both on the relationship of each member to the head of the household and on the presence of pointer variables in the original microdata, such as those in IPUMS-I and LFS, which identify familial relationships derived from the household roster. These pointer variables are crucial for linking individuals within the same household and for understanding familial structures. Our algorithm specifically uses pointer variables that indicate the line number of a person’s mother (MOMLOC in IPUMS and HHMOTH in LFS), father (POPLOC and HHFATH), or spouse (SPLOC and HHSPOU). For enhanced robustness, consistency, and accuracy, the algorithm also incorporates the NCHILD variable from IPUMS-I, which we have implemented in LFS data where it was previously absent. This variable counts the number of a person's own children living in the household. As noted, the algorithm generates a living arrangement value that reflects the co-residence with up to nine different relationships to the ego, calculated using powers of two. The living arrangement index ranges from 0, indicating that a person lives alone, to 511, representing a case where an individual lives with their father, mother, child, partner, at least one sibling, grandparent, grandchild, other relative, and non-relative. We define living with "other relative" as someone who shares a household with a relative who is not their father, mother, child, sibling, grandparent, or grandchild. To illustrate the algorithm's functionality, Table 1 presents its application to a hypothetical multigenerational household consisting of 13 individuals.

RELATED	PERNUM	MOMLOC	POPLOC	SPLOC	LAI	LAI qualitative
Head	1	10	11	2	463	Father+Mother+Child+Partner+Grandchild+Other relative+Non-relative
Partner	2	12	0	1	462	Mother+Child+Partner+Grandchild+Other relative+Non-relative
Child	3	2	1	0	435	Father+Mother+Sibling+Grandaparent+Other relative+Non-relative
Child	4	2	1	0	435	Father+Mother+Sibling+Grandaparent+Other relative+Non-relative
Child	5	2	1	6	383	Father+Mother+Child+Partner+Sibling+Grandaparent+Grandchild+Non-relative
Other relative	6	0	0	5	460	Child+Partner+Grandchild+Other relative+Non-relative
Grandchild	7	5	6	0	439	Father+Mother+Child+Sibling+Grandaparent+Other relative+Non-relative
Grandchild	8	5	6	0	435	Father+Mother+Sibling+Grandaparent+Other relative+Non-relative
Non relative	9	0	0	0	256	Non-relative
Parent	10	0	0	11	460	Child+Partner+Grandchild+Other relative+Non-relative
Parent	11	0	0	10	460	Child+Partner+Grandchild+Other relative+Non-relative
Parent	12	0	0	0	452	Child+Grandchild+Other relative+Non-relative
Other relative	13	7	0	0	418	Mother+Grandaparent+Other relative+Non-relative

Table 1: Implementation of the algorithm to reconstruct living arrangements

One advantage of our algorithm and classification method is its flexibility, allowing the resulting categories to be grouped in various ways to create typologies of living arrangements tailored to the specific needs and goals of different research studies. In the database presented here the 512 potential combinations that the living arrangements index yields have been grouped into 8 categories and 11 subcategories that answer the question "With whom does a given ego live?" as shown in Table 2. The 8 main categories consist of:

1. **Living Alone:** ego lives in a unipersonal household.

2. **With single parent:** ego lives with a single parent or with a single parent and his/her siblings.

2.1 **With single parent extended:** ego lives with a single parent and other relatives or with a single parent, his/her siblings and other relatives.

2.2 **With single parent extended composite:** ego lives with a single parent, other relatives (excluding siblings) and at least one non-relative, or with a single parent, his/her siblings, other relatives and at least one non-relative.

3. **With parents:** ego lives with both parents, or with both parents and his/her siblings.

3.1 **With parents extended:** ego lives with both parents and other relatives (excluding siblings), or with both parents, his/her siblings and other relatives.

3.2 **With parents extended composite:** ego lives with both parents, other relatives (excluding siblings) and at least one non-relative, or with both parents, his/her siblings, other relatives and at least one non-relative.

4. **With partner:** ego lives with his/her partner.

4.1 **With partner extended:** ego lives with his/her partner and other relatives (excluding parents).

4.2 **With partner extended composite:** ego lives with his/her partner, other relatives (excluding parents) and at least one non-relative.

5. **With partner and children:** ego lives with his/her partner and children.

5.1 **With partner and children extended:** ego lives with his/her partner, children and other relatives.

5.2 **With partner and children extended composite:** ego lives with his/her partner, children, other relatives and at least one non-relative.

6. **With children:** ego lives with children and no partner.

6.1 **With children extended:** ego lives with children, no partner and other relatives.

6.2 **With children extended composite:** ego lives with children, no partner, other relatives and at least one non-relative.

7. **With extended family:** ego lives with other relatives other than his/her parents or children.

7.1 **With extended family composite:** ego lives with other relatives other than his/her parents or children and at least one non-relative.

8. **With non-relatives:** ego lives exclusively with non-relatives.

Living Arrangement Index (LAI)	Living Arrangement Index qualitative
0	Living alone
1	Father
2	Mother
3	Father+Mother
4	Child
5	Father+Child
6	Mother+Child
7	Father+Mother+Child
8	Partner
9	Father+Partner
10	Mother+Partner
11	Father+Mother+Partner
12	Child+Partner
13	Father+Child+Partner
14	Mother+Child+Partner
15	Father+Mother+Child+Partner
16	Sibling
17	Father+Sibling
18	Mother+Sibling
19	Father+Mother+Sibling
20	Child+Sibling
21	Father+Child+Sibling
22	Mother+Child+Sibling
23	Father+Mother+Child+Sibling
24	Partner+Sibling
25	Father+Partner+Sibling
26	Mother+Partner+Sibling
27	Father+Mother+Partner+Sibling
...	
511	Father+Mother+Child+Partner+Sibling+ Grandaparent+Grandchild+Other relative+Non-relative

Living Arrangement type	Living Arrangement type qualitative
10	**Living Alone**
20	**With single parent**
21	With single parent extended
22	With single parent extended composite
30	**With parents**
31	With parents extended
32	With parents extended composite
40	**With partner**
41	With partner extended
42	With partner extended composite
50	**With partner and children**
51	With partner and children extended
52	With partner and children extended composite
60	**With children**
61	With children extended
62	With children extended composite
70	**Extended**
71	Extended composite
80	**Non-relative**

Table 2: Aggregation of Living arrangement Index into Living Arrangement types**.** The full table of conversion can be accessed here: https://github.com/JuanGaleano/CORESIDENCE/blob/main/AGGREGARION_LAI_LAT.csv

1.5 Harmonization processes

To construct the Harmonized dataset within CoLADB, the process began by aggregating the single-age data provided in most IPUMS-I samples into five-year age groups, as LFS data is disseminated. Following this, it was necessary to harmonize the categories related to educational attainment and marital status across samples to ensure consistency and comparability between data sources. Both the EU-LFS and IPUMS International utilize the International Standard Classification of Education (ISCED) levels to categorize educational attainment. However, because IPUMS samples often include more detailed categories, we collapsed these categories to align with the broader classifications used in the EU-LFS. The following table depicts this procedure. In the case of the marital status categories recorder in each source, the only modification needed in IPUMS-I samples was to aggregate the category widowed to that of divorced and separated.

IPUMS-I	EU-LFS
1. Less than primary completed 2. Primary completed	1. Low Education: Less than primary, primary and lower secondary (ISCED levels 0-2)
3. Secondary completed	2. Medium Education: Upper secondary and post-secondary non-tertiary (ISCED levels 3 and 4)
4. University completed	3. High Education: Short-cycle tertiary, bachelor or equivalent, master or equivalent and doctoral or equivalent (levels 5-8)

Table3: Original educational categories in IPUMS-I and EU-LFS

2. Data records

The CoLADB is hosted in Zenodo, an open-access digital repository that allows researchers, scientists, and scholars from various disciplines to share and preserve their research outputs. Zenodo is operated by CERN (European Organization for Nuclear Research) and supported by various organizations, including the European Commission's OpenAIRE project. The CoLADB is hosted at the permanent DOI: https://doi.org/10.5281/zenodo.8142652. The repository is composed of the following elements: a RData file named CORESIDENCE_LA_DB containing the CoLADB in the form of a List. In R, a List object is a versatile data structure that can contain a collection of different data types, including vectors, matrices, data frames, other lists, spatial objects or even functions. It allows to store and organize heterogeneous data elements within a single object. The CORESIDENCE_LA_DB R-list object is composed of six elements:

1. IPUMS: a data frame where data is aggregated by single ages, marital status, educational attainment and living arrangement types. Source of the original data: IPUMS-I
2. IPUMSAG: a data frame where data is aggregated by five-year age groups, marital status, educational attainment and living arrangement types. Source of the original data: IPUMS-I
3. LFS: a data frame where data is aggregated by five-year age groups, marital status, educational attainment and living arrangement types. Source of the original data: EU-LFS.
4. HARMONIZED: a data frame where data is aggregated by five-year age groups, marital status, educational attainment and living arrangement types. The categories of marital status and educational attainment have been harmonized between the two data sources. Source of the original data: IPUMS-I and EU-LFS
5. CODEBOOK: a data frame with the complete list of indicators, their code names and description.

3. Technical Validation

As outlined in the Background and Summary section, global data for analysing the composition, distribution, and evolution of living arrangements over time is notably limited. This scarcity makes it difficult to externally validate the innovative data presented by CoLADB. To address this, we rely on the IPUMS-I samples for validation. Specifically, we utilize the constructed variable HHTYPE (household type), which defines 11 types of households. Of these, four categories—unipersonal households, married couples without children, married couples with children, and single-parent households— are directly comparable with the living arrangement types in CoLADB. We compare the share of population within each sample living in the different household types of IPUMS-I against the share of population within each sample in the equivalent living arrangements types of CoLABD. By leveraging HHTYPE, we assess the consistency and accuracy of our algorithm's results against this established source, ensuring the reliability and validity of our data.

Overall, the correlation between the country-level indicator of the CoLADB and the ones from the IPUMS-I is highly linear, suggesting a very good fit of our computations (Figure 3). Additionally, we computed an equal variance T-test for each of the selected indicators. The p-values, greater than the common significance level of 0.05, suggest that the observed difference in means is likely due to random variation, primarily associated with the data cleaning and processing steps. This indicates that the disparities between the compared databases are more likely a result of data handling rather than genuine differences in means.

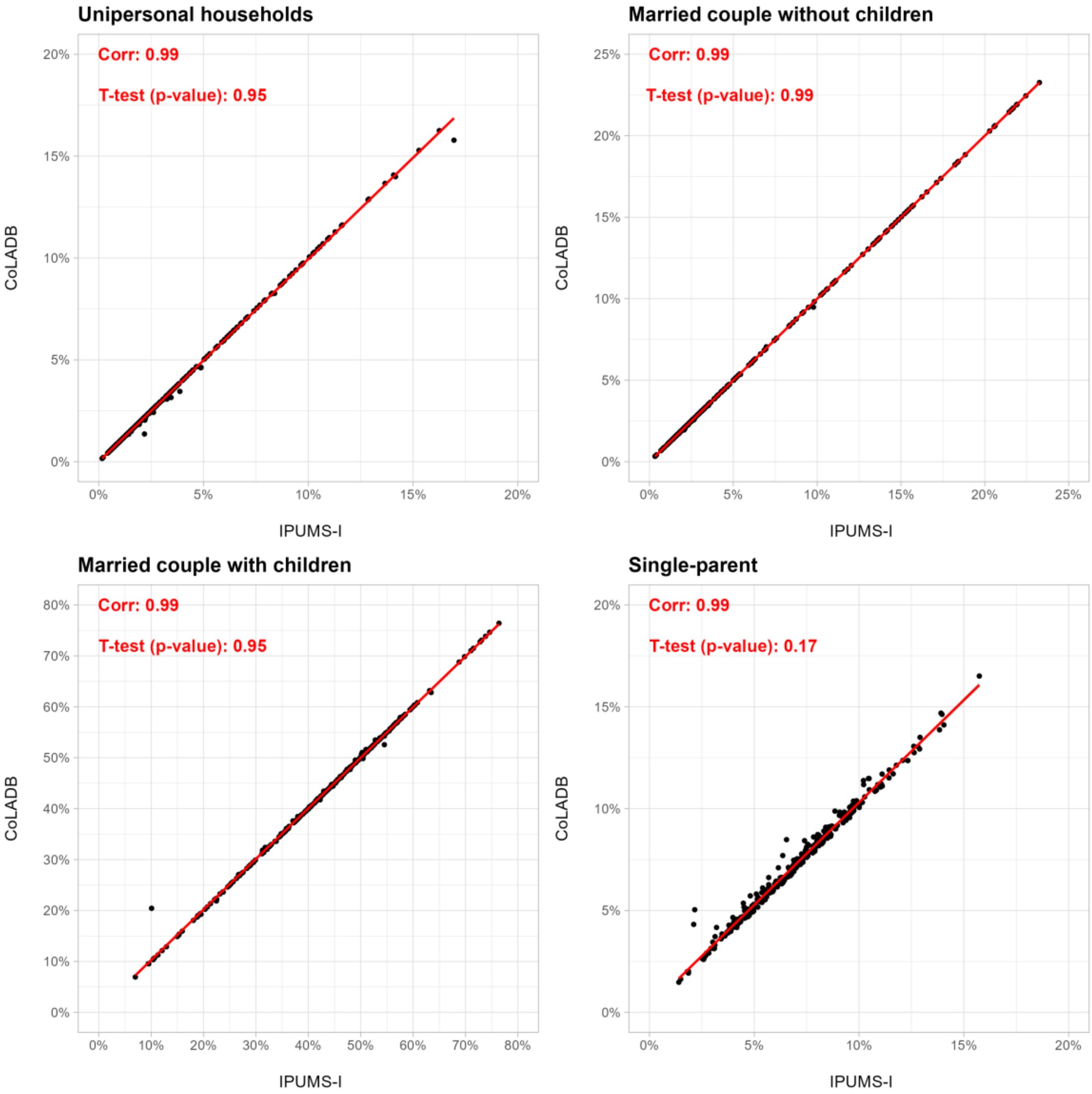

Fig. 5: Validation of the CoLADB

4. Code availability

The processing steps to build the three datasets composing the CoDB were carried out in R, utilizing the libraries tidyerse[18], haven[19], labelled[20], tibble[21] and parallel[15]. All the code is available on the GitHub repository of this project: https://github.com/JuanGaleano/CORESIDENCE

5. Acknowledgements

We thank Guillem Esteve for his initial proposal for computing the living arrangement index implemented for the creation of this database.

6. Funding

European Research Council (ERC). Advanced Grant. Reference: HE-ERC-2021-AdG-GA No 101052787-CORESIDENCE

7. Author contributions

Albert Esteve conceived the project, is the principal investigator of the CORESIDENCE project and reviewed the initial manuscript.

Juan Galeano, designed the analytic strategy and processed the data, wrote the R code for building the CoLADB, and wrote the initial manuscript and prepared the figures for it.

8. Competing interests

The authors declare no competing interests.

9. References

[1] Furstenberg, Frank F. *Destinies of the Disadvantaged: The Politics of Teen Childbearing.* Russell Sage Foundation, (2007). https://www.jstor.org/stable/10.7758/9781610442343

[2] Elder, G. H. Jr. "The Life Course as Developmental Theory." Child Development, 69(1), 1-12, (1998) https://doi.org/10.2307/1132065

[3] Furstenberg, F. F. Banking on families: How families generate and distribute social capital. Journal of Marriage and Family, 67(4), 809–821. (2005). https://doi.org/10.1111/j.1741-3737.2005.00177.x

[4] Bourdieu, P. "The Forms of Capital." In J. G. Richardson (Ed.), Handbook of Theory and Research for the Sociology of Education (pp. 241-258). Greenwood Press. (1986)

[5] Giddens, A. (1991). Modernity and Self-Identity: Self and Society in the Late Modern Age. Stanford University Press. http://www.sup.org/books/title/?id=2660

[6] Lesthaeghe, R. The unfolding story of the second demographic transition. Population and Development Review, 36(2), 211-251. (2010) https://doi.org/10.1111/j.1728-4457.2010.00328.x

[7] McLanahan, S., & Percheski, C. Family structure and the reproduction of inequalities. Annual Review of Sociology, 34, 257-276. (2008). https://doi.org/10.1146/annurev.soc.34.040507.134549

[8] Esping-Andersen, G. *Social foundations of postindustrial economies.* Oxford University Press. (1999) https://doi.org/10.1093/0198742002.001.0001

[9] Bongaarts, J. "Fertility Transitions in Developing Countries: Progress or Stagnation?" Studies in Family Planning, 39(2), 105-110. (2008). https://doi.org/10.1111/j.1728-4465.2008.00157.x

[10] Lutz, W., & Samir, K. C. "Dimensions of Global Population Projections: What Do We Know About Future Population Trends and Structures?" Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1554), 2779-2791. (2010). https://doi.org/10.1098/rstb.2010.0133

[11] United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2019: Highlights. New York: United Nations. (2019). https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf

[12] Oeppen, J., & Vaupel, J. W. Broken limits to life expectancy. Science, 296(5570), 1029-1031. (2002). https://doi.org/10.1126/science.1069675

[13] Alburez-Gutierrez, Diego, Williams, Ivan & Caswell, Hall. Projections of human kinship for all countries. Proceeding of the National Academy of Science, vol. 120 (52) (2023). https://doi.org/10.1073/pnas.2315722120

[14] Galeano, J., Esteve, A., Turu, A. et al. CORESIDENCE: National and subnational data on household size and composition around the world, 1964–2021. Sci Data 11, 145 (2024). https://doi.org/10.1038/s41597-024-02964-3

[15] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (2024).

[16] Minnesota Population Center. IPUMS International [multiple datasets]. Minneapolis, MN: University of Minnesota. (2023).

[17] Eurostat. European Union Labour Force Survey [multiple datasets]. European Union. (2023

[18] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... & Hester, J. Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686 (2019)

[19] Wickham, H., & Miller, E. haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.3.1. https://CRAN.R-project.org/package=haven (2020).

[20] Christoph, J., & Lazarevic, M. labelled: Manipulating Labelled Data. R package version 2.8.0. https://CRAN.R-project.org/package=labelled (2020).

[21] Müller, K., & Wickham, H. tibble: Simple Data Frames. R package version 3.1.4. https://CRAN.R-project.org/package=tibble (2021).