Artículo de revista:

Galeano, Juan & Esteve, Albert (2025). “The Global Living Arrangements Database, 1960–2021”. *Scientific Data* **12**:1483 (ISSN 2052-4463) https://doi.org/10.1038/s41597-025-05787-y scientific data

OPEN

DATA DESCRIPTOR

The Global Living Arrangements Database, 1960–2021

Juan Galeano ✉ & Albert Esteve

This paper introduces the Global Living Arrangements Database (GLAD), a global resource designed to fill a critical gap in the availability of statistical information for examining patterns and changes in living arrangements by age, sex, marital status and educational attainment. Utilizing comprehensive census microdata from IPUMS International and the European Labour Force Survey (EU-LFS), GLAD summarizes over 740 million individual records across 107 countries, covering the period from 1960 to 2021. This database has been constructed using an innovative algorithm that reconstructs kinship relationships among all household members, providing a robust and scalable methodology for studying living arrangements. GLAD is expected to be a valuable resource for both researchers and policymakers, supporting evidence-based decision-making in areas such as housing, social services, and healthcare, as well as offering insights into long-term transformations in family structures. The open-source R code used in this project is publicly available, promoting transparency and enabling the creation of new ego-centred typologies based in interfamily relationships.

Background & Summary

Lives are shaped by moments shared with others or spent in solitude. Regardless of the circumstances, many of these moments unfold in the context of households. The people individuals live with—or their absence—profoundly influence who they are and the opportunities available to them[1]. At every stage of life, roles as parents, children, siblings, partners, friends, or roommates—as well as personalities and well-being—are intricately tied to living arrangements[2,3]. These arrangements not only affect individuals but also reflect and reinforce societal norms that shape how communities are organized[4–6].

Across societies, most people live in private households, with average household sizes ranging from 2 to 8 members, depending on the country[7]. These households are typically composed of first-degree relatives or other family members, although individuals may also live alone or with non-relatives[8]. Despite these broad patterns, the structure and composition of living arrangements vary widely across time and place. In some contexts, households are centred around the nuclear family, while in others, multiple generations share the same home, reflecting the vast heterogeneity of human experience.

The implications of living arrangements—ranging from intra-household care and emotional well-being to gender roles and consumption—have been widely researched[9,10]. Historically, research on these arrangements has prioritized the household perspective, classifying households by size, structure, and typology[11]. Several databases provide data on households at the international level. The United Nations began publishing global household data in 2022, covering 196 countries[12]. More recently, the "Intergenerational Coresidence in Global Perspective: Dimensions of Change (CORESIDENCE)" project, funded by the European Research Council, contributed to this effort by releasing the CoDB database[13]. This database includes additional indicators and harmonized data at the subnational level, offering a more nuanced understanding of household composition and its evolution over time. The household has long been used as the primary unit of enumeration in censuses and surveys. Information about individuals residing in the same dwelling is typically collected through a standardized questionnaire known as the household roster. Among other limitations, connected with a western-centric definition of households, one major analytical shortcoming of the household roster is its unidirectional classification of relationships[14]—defining all household members in relation to a single individual designated as the head or reference person and therefore limiting a nuance analysis of living arrangements.

Continuing on previous efforts, the CORESIDENCE project now introduces the Global Living Arrangement Database (GLAD). GLAD offers several innovations. First, it is the world's first global database on living arrangements analysed from an individual-based perspective, enabling disaggregation by age, sex, education level, and marital status. Second, it introduces a globally comparable, yet flexible, typology of living arrangements,

Centre d'Estudis Demogràfics / Universitat Autònoma de Barcelona, Barcelona, Spain. ✉e-mail: jgaleano@ced.uab.es

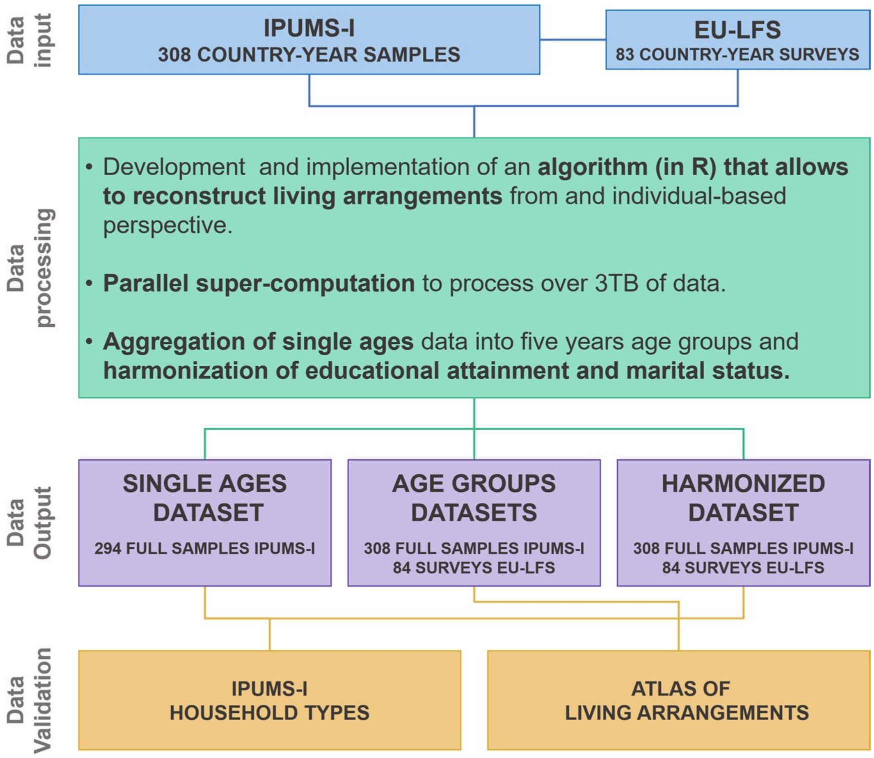

Fig. 1 Flowchart representing the different stages to build the GLAD.

allowing for reclassification based on specific research needs. Third, it draws on data from over 740 million individual records collected through censuses and surveys across 107 countries between 1960 and 2020. Each country-year sample included allows to group individuals into households and facilitates the analysis of kin and non-kin intra-household relationships. All together, these features make GLAD a valuable tool for analysing variations in living arrangements across age, sex, educational level, and marital status, while also enabling comparisons across different contexts and temporal dimensions

To ensure transparency and replicability, the open-source R code developed to produce GLAD is publicly available. This code enables researchers to scrutinize how the microdata was processed and how the typology of living arrangements was constructed. It also empowers users to create new classifications tailored to their specific research goals.

GLAD opens up vast research possibilities. It allows for analyses of living arrangements from a new individual-based perspective by age, sex, educational attainment, and marital status across different countries and time periods, overcoming the limitation of the household roster[15]. It addresses fundamental questions about the organization of human societies, such as: With whom do we share our lives at home? Which are the most common living arrangements? How do they vary by sex and age in different societies? What characteristics define households at different stages of life? How do extended versus nuclear forms of living arrangements balance over the life course in different countries? How do gender differences in union formation and dissolution shape distinct living arrangements? Beyond these micro-level inquiries, GLAD also facilitates macro-level analyses, such as exploring differences in living arrangements between high- and low-income countries or understanding how demographic changes drive transformations in household composition.

Methods

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

GLAD 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) repository (see section **Methods: Data Sources**).

All data cleaning, processing, harmonization, and aggregation were conducted using R[16]. The complete code for constructing GLAD is available in the project's GitHub repository (see section **Code Availability**). The vast scale of this project—encompassing over 785 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 use the processing capacity of the Barcelona Supercomputing Centre (BSC), home to MareNostrum V (https://www.bsc.es/). This collaboration, which

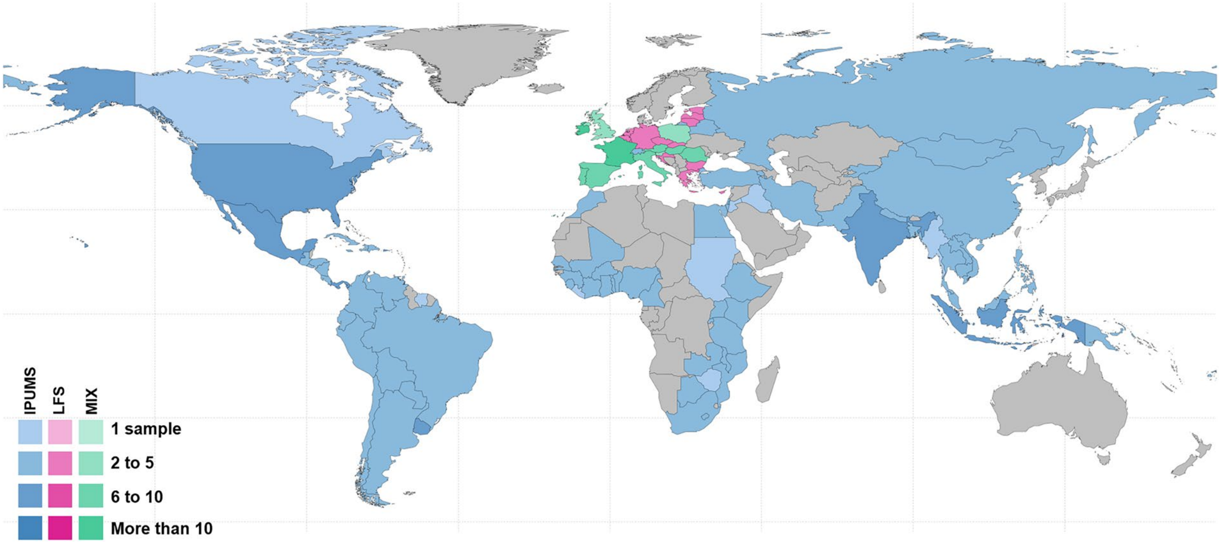

Fig. 2 Country coverage by number of samples available of the GLAD.

bridges Social and Computational Sciences by integrating computer engineering, parallel computing, and demographic analysis, has enabled us to process the extensive data underlying GLAD.

The output data of GLAD consists of four distinct datasets based on the source of the microdata, the disaggregation by age and the harmonization of educational level and marital status.

(1) Living Arrangements (LA) by Single Age (based on IPUMS data): This dataset includes 301 country-year samples from 91 countries worldwide (Fig. 2), covering over 760 million individual records from the full samples available in IPUMS-I. The original microdata is aggregated by single age, educational attainment, and marital status—using IPUMS-I's defined categories—as well as by set of types of living arrangements (see Section **Methods: Living Arrangements**).

(2-3) Living Arrangements by Five-Year Age Groups (based on IPUMS and LFS data): These datasets, presented separately, are derived from 319 country-year samples from IPUMS-I (2) and 86 from LFS (3), encompassing a total of 787 million individual records across 107 countries worldwide. They retain the original categories for educational attainment and marital status as defined by their respective sources, which are detailed in the GLAD codebook (see Section **Data Records**).

(4) Living Arrangements by Five-Year Age Groups with Harmonized Educational Level and Marital Status (IPUMS and LFS Combined): This dataset aggregates information into five-year age groups (using single-age IPUMS-I samples) and classifies individuals by living arrangement type, marital status, and educational attainment. To enable comparisons across samples from different sources, the categories for educational attainment and marital status have been harmonized (see Section **Methods: Harmonization Process**).

GLAD offers near-universal coverage, representing countries from all continents (Fig. 3). However, compared to our previous household-level database (CoDB)[13], it includes fewer samples. This difference arises because not all censuses and surveys samples in GLAD provide the necessary information to reconstruct relationships among household members. While samples from IPUMS and the European Labour Force Surveys (EU-LFS) include relationship variables—also known as pointer variables—that enable the application of our algorithm for reconstructing living arrangements, others, such as the ones from the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS), do not. The absence of these variables makes reconstructing household relationships significantly more complex and challenging as there is not a consistent way to link children to mothers in complex households.

To ensure the accuracy and reliability of the GLAD, 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. We also provide a leaflet with three descriptive graphs for each sample to ensure transparency regarding both the microdata underlying GLAD and the results produced by our algorithm for reconstructing living arrangements from an individual-based perspective (see section **Technical Validation**).

Data sources. The GLAD 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.

The primary source of individual microdata for GLAD is the International Integrated Public Use Microdata Series (IPUMS-I)[17], which consists of 308 census samples from 94 countries. The IPUMS International project is renown global initiative dedicated to the collection, preservation, harmonization, and distribution of census microdata from across the globe[18]. For building GLAD, we utilized the complete samples available in IPUMS-I, encompassing more than 720 million individual observations. This extensive dataset provides a robust foundation for analysing living arrangements trends at a global scale. It's worth mentioning in the case of the IPUMS-I samples, which were derived mainly from population censuses, we only kept and provide information

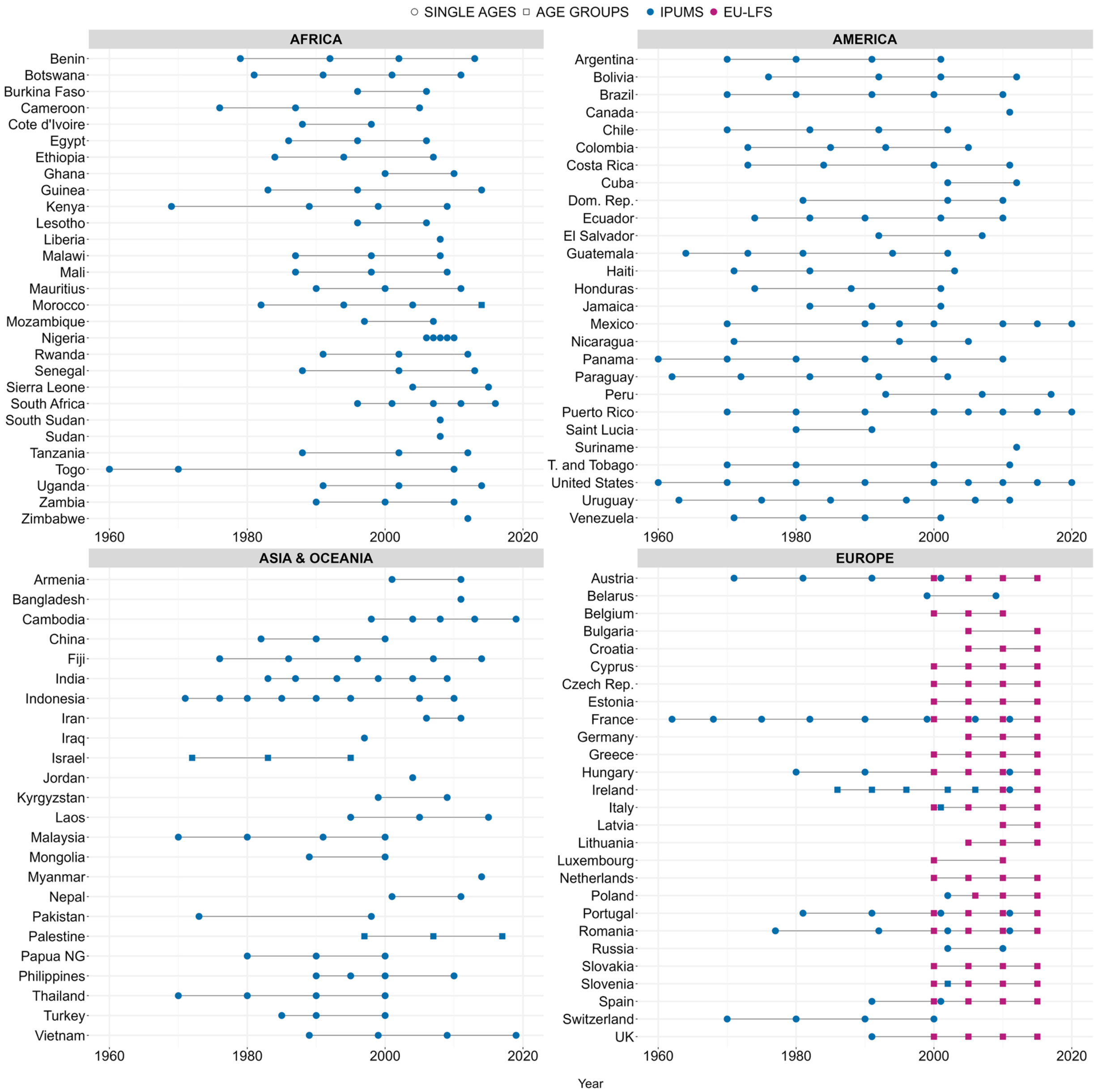

Fig. 3 Availability of samples by country, year, ages and source in the GLAD.

for individuals living in private households, dropping out of the samples those people who lived in groups quarters, because they cannot consistently identify across samples.

In order to complement the scarce information available on European countries from IPUMS-I, we processed, as secondary source of individual microdata, 86 samples of the European Labour Force Survey (EU-LFS)[19]. 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, or not, among them. Crucially for our porpoises, samples from year 2000 onwards included the necessary pointer variables for the reconstruction of living arrangements. As EU-LFS collects data on a quarterly basis, samples included in the GLAD correspond to the yearly samples to ensure consistency with the specific time frame for which the data was downloaded.

The GLAD is an aggregated database that does not contain individualized information. As a secondary data source, the quality of the information it provides is inherently linked to that of the original census microdata from which it is derived (see section **Technical Validation: Atlas of Living Arrangements**). One important limitation is that the reconstruction of living arrangements implemented in this project does not allow us to systematically distinguish between arrangements resulting from a previous marriage dissolution and those that are not. For instance, in a stepfamily consisting of two adults and two children, each adult is classified as living *with partner and children*, while each child is classified as living *with parents and sibling*, without any indication of step-relationships within the household. Despite this limitation, the GLAD represents a major contribution to the study of living arrangements worldwide. It provides harmonized and systematically classified data across a vast number of countries and time points, enabling consistent cross-national and over-time comparisons. While certain nuances—such as step-relationships—are inevitably lost in the process of data harmonization and aggregation, the broad patterns and typologies captured by GLAD offer critical insights into global family and household transformations that would otherwise remain inaccessible at this scale.

Designed with a forward-looking perspective, the GLAD is poised to accommodate the ongoing growth of its constituent data repositories. As these underlying data sources continue to release new samples, the GLAD is primed to seamlessly integrate these additions, ensuring its continued comprehensiveness and relevance over time.

Living arrangements. A core objective in constructing the database has been to classify the population according to their living arrangements. To achieve this objective, we followed a three-step process: First, we developed a comprehensive classification of living arrangements, referred to here as Living Arrangements Values (LAV), which we applied to the microdata on which GLAD is built. Second, the data are aggregated using a numerical code consisting of two digits denoting the Living Arrangements Type (LAT) and seven digits representing the Living Arrangements Index. This coding system enables data aggregation while preserving a high level of detail.

Living arrangement values (LAV). A key contribution to the creation of GLAD was the development of an algorithm that reconstructs kinship relationships among household members by assigning a living arrangement value to each individual. The algorithm leverages both the reported relationship to the head of the household and the presence of pointer variables in the original microdata—such as those in IPUMS-I[20] and EU-LFS—which identify familial links derived from the household roster. These pointer variables were essential for connecting individuals within the same household and understanding family structures.

The pointer variables used by the algorithm indicate the line number of a person's mother (MOMLOC in IPUMS, HHMOTH in EU-LFS), father (POPLOC in IPUMS, HHFATH in EU-LFS), and spouse (SPLOC in IPUMS, HHSPOU in EU-LFS). To enhance robustness, it also incorporated the NCHILD variable from IPUMS-I, which we implemented in EU-LFS data where it was previously absent. NCHILD counts the number of a person's own children living in the household. To ensure accuracy, the algorithm processed each sample at the household level. Before computation, samples were split into their constituent private households, allowing precise reconstruction of family relationships within each unit avoiding miscomputations. The code of the algorithm is available in the project's GitHub repository (see section **Code Availability**) and it can be defined as follows:

$$LAV = \sum_{i=1}^{9} v_i 2^{i-1}$$

$$v = (v_1, \ldots, v_i, \ldots v_9)$$

$$v_i = \begin{cases} 0 = Abscense \\ 1 = Presence \end{cases}$$

Where v_i denotes the nine possible relation to a given ego. In qualitative terms, the algorithm operates as a numerical encoding system that detects kinship based on nine potential relationships to the reference individual or "ego": *father*, *mother, child, partner, sibling, grandparent, grandchild, other relative*, or *non-relative*. These relationships were encoded into a living arrangement value (LAV) by summing whether each of ego's nine potential relationships was present in the household. Each relationship was represented using powers of two, ranging from 2^0 to denote the presence of ego's father in the household to 2^8 for cohabitation with a non-relative. Table 1 illustrates how this numerical encoding system works by presenting the case of a person who shared a household with three generations. This individual lived with his or her mother, at least one child, a partner, at least one grandparent, and at least one other relative, scoring a living arrangement value of 174. All possible values (512 in total) correspond to a unique combination of relatives and no-relatives in the household.

The living arrangement values range from 0, indicating that a person lives alone, to 511, representing the hypothetical, although unlikely, case where an individual lives with his or her father, mother, child, partner, at least one sibling, grandparent, grandchild, other relative, and non-relative. We define living with "other relative" as someone who shared a household with a relative who is not their father, mother, child, sibling, grandparent, or grandchild.

The algorithm is flexible, allowing the resulting values to be grouped in various ways to create typologies of living arrangements tailored to the specific needs and goals of different research studies. Table 2 presents the application of the algorithm to a hypothetical multigenerational household consisting of 13 individuals.

Living arrangement types (LAT). In the database presented here the 512 potential combinations that the living arrangements algorithm yields are aggregated into a Living Arrangement Index (LAI) composed of 9 digits. The first two digits of the string denote 8 Living Arrangement Type (LAT) and 11 subtypes that answer the question "With whom does a given ego live?" as shown in Table 3.

Relation to ego	Code relation to ego	Value	Presence (1)/absence (0)	LAV
Father	1	2^0 = 1	0	0
Mother	2	2^1 = 2	1	2
Child	3	2^2 = 4	1	4
Partner	4	2^3 = 8	1	8
Sibling	5	2^4 = 16	0	0
Grandparent	6	2^5 = 32	1	32
Grandchild	7	2^6 = 64	0	0
Other relative	8	2^7 = 128	1	128
Non-relative	9	2^8 = 256	0	0
Total		511		174

Table 1. Numerical encoding system used to assign the corresponding Living Arrangement Value (LAV) to each individual, with application to a concrete example.

The proposal of the Living Arrangements Types (LAT) is logically revisable and can be modified by researchers based on the code we provide over the microdata. However, based on the exploration of the results performed by the CORESIDENCE team, this typology adapts to different countries, encompasses groups with a certain number of cases, and links to previous household classifications. The typology is based on first-degree relatives (parents, children, and partner), who constitute what is referred to as the family nucleus. Once family nuclei are identified and each individual's position within the nucleus is determined, the presence of other relatives defines the *extended* nature of the living arrangement, while the presence of non-relatives defines its *composite* character. The classification begins with the simplest and smallest household type—living alone—and gradually incorporates primary-kin relatives, prioritizing the formation voluntary agency of a nucleus by ego in the classification and understanding by "voluntary agency" when ego has either a partner and/or a child. For example, if a person lives with their partner and at least one parent, they are classified as "With partner extended." Similarly, if a person lives with at least one child, without a partner, and with at least one parent, they are categorized as "With children extended." Following this logic, if a person lives with one parent a grandparent a grandchild and a non-relative or with a parent a grandparent a grandchild, one other relative (as it could be an uncle or aunt) and a non-relative, in both cases these individuals are categorized al living "With single parent extended composite". The suffix "composite" in this classification indicates the presence of a household member with whom the ego under evaluation has no kin relationship.

The 8 main living arrangements types and their subtypes are defined as follows and illustrated visually in Fig. 4:

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.
    4.2 **With partner extended composite:** ego lives with his/her partner, other relatives 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.

RELATED	PERNUM	MOMLOC	POPLOC	SPLOC	LAV	Living Arrangement Value 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 + Grandparent + Other relative + Non-relative
Child	4	2	1	0	435	Father + Mother + Sibling + Grandparent + Other relative + Non-relative
Child	5	2	1	6	383	Father + Mother + Child + Partner + Sibling + Grandparent + 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 + Grandparent + Other relative + Non-relative
Grandchild	8	5	6	0	435	Father + Mother + Sibling + Grandparent + 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 + Grandparent + Other relative + Non-relative

Table 2. Implementation of the algorithm to reconstruct living arrangements.

7. **With other extended:** ego lives with other relatives other than his/her parents or children.
   7.1 **With other extended 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). After the first two digits, which denote the Living Arrangement Type as explained above, the remaining seven digits specify the individual's living arrangement based on nine possible relationships to ego. Specifically, the third digit indicates the presence or absence of the individual's father, the fourth digit represents the mother, and the fifth digit corresponds to siblings. Similarly, the sixth digit denotes the presence of grandchildren, the seventh represents grandparents, the eighth accounts for other relatives, and the ninth and final digit indicates whether the individual lives with a non-relative. The possible value of each digit of the string composing the living arrangement index is indicated in Fig. 5.

The living arrangement index enables the decomposition of living arrangement types. For example, a row in GLAD with a LAI value of 321230567 represents the population of a given age, sex, marital status, and educational attainment in a living arrangement of type 32 (With parents, extended composite). This index indicates that these individuals shared a household with their father and mother, at least one sibling, at least one grandparent, at least one other relative (such as an uncle, aunt, a cousin or a child in law), and at least one person with whom they have no kin relationship.

Additional variables available: marital status and educational attainment. To construct the Harmonized dataset within GLAD, 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. Table 4 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 Table 4.

Data Records

The GLAD is hosted in Zenodo, at the permanent https://doi.org/10.5281/zenodo.15038209[21]. The repository is composed of the following elements: a Rda file named CORESIDENCE_GLAD 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. All datasets that comprise GLAD, together with the codebook, are also available as CSV files in the same repository. The CORESIDENCE_GLAD R-list object is composed of six elements:

1. SINGLE AGES: a data frame where data is aggregated by single ages, marital status, educational attainment and living arrangement types. Source of the original data and number of samples: IPUMS-I, 294 samples.
2. AGE GROUPS IPUMS: 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 and number of samples: IPUMS-I, 304 samples.

LAV	Living Arrangement Value	LAT	Living Arrangement Type
0	Living alone	10	Alone
1	Father	20	With single parent
2	Mother	20	With single parent
3	Father + Mother	30	With parents
4	Child	60	With children
5	Father + Child	61	With children extended
6	Mother + Child	61	With children extended
7	Father + Mother + Child	61	With children extended
8	Partner	40	With partner
9	Father + Partner	41	With partner extended
10	Mother + Partner	41	With partner extended
11	Father + Mother + Partner	41	With partner extended
12	Child + Partner	50	With partner and children
13	Father + Child + Partner	51	With partner and children extended
14	Mother + Child + Partner	51	With partner and children extended
15	Father + Mother + Child + Partner	51	With partner and children extended
16	Sibling	70	Extended
17	Father + Sibling	20	With single parent
18	Mother + Sibling	20	With single parent
19	Father + Mother + Sibling	30	With parents
20	Child + Sibling	61	With children extended
21	Father + Child + Sibling	61	With children extended
22	Mother + Child + Sibling	61	With children extended
23	Father + Mother + Child + Sibling	61	With children extended
24	Partner + Sibling	41	With partner extended
25	Father + Partner + Sibling	41	With partner extended
26	Mother + Partner + Sibling	41	With partner extended
27	Father + Mother + Partner + Sibling	41	With partner extended
…		…	
511	Father + Mother + Child + Partner + Sibling + Grandparent + Grandchild + Other relative + Non-relative	52	With partner and children extended composite

Table 3. Aggregation of Living arrangement Index into Living Arrangement types. The full table of conversion can be accessed in the GitHub page of the project, see section **Code Availability**.

3. AGE GROUPS 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 and number of samples: EU-LFS, 84 samples.
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 and number of samples: IPUMS-I and EU-LFS, 391 samples.
5. CODEBOOK: a data frame with the complete list of variables included, their names description and categories.
6. LABELS LAT: A R function to add the qualitative labels to Living Arrangement Types (LAT).
7. ATLAS LIVING ARRANGEMENTS: The url of the folder with leaflet of living arrangements for each sample included in GLAD.

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, which makes it difficult to externally validate the innovative data presented by GLAD. To address this, we rely on the 294 IPUMS-I samples included in the SINGLE AGES dataset for validation. Together, these samples comprise over 710 million unweighted individual cases—accounting for 96.6% of the total unweighted individual cases included in GLAD. 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

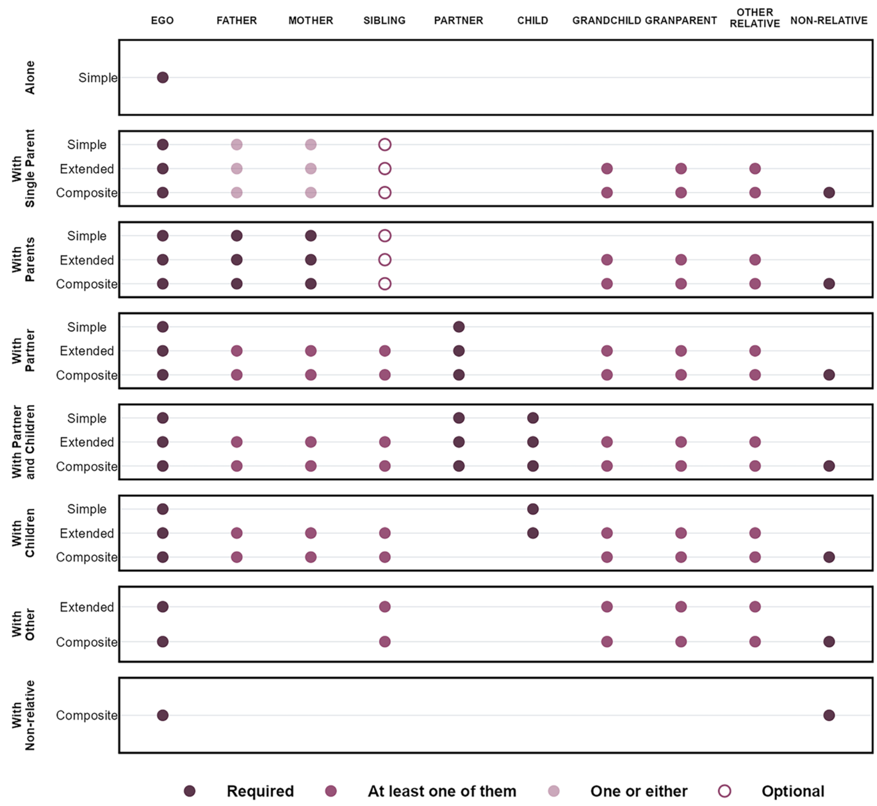

Fig. 4 Configuration of Living Arrangement Types (LAT).

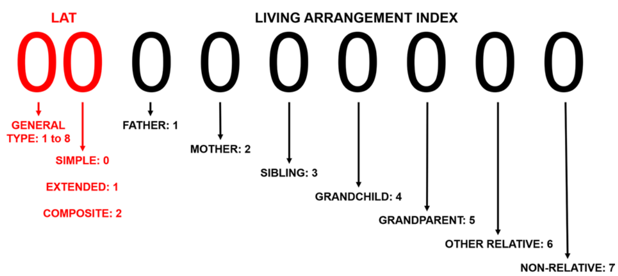

Fig. 5 Value of each digit of the string composing the living arrangement index.

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)

Table 4. Original educational categories in IPUMS-I and EU-LFS.

types in GLAD. 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 GLAD. The comparison with HHTYPE serves not as a definitive external validation but rather as a consistency check. It helps demonstrate that, at an aggregate level, our classification does not significantly diverge from the well-established IPUMS derivations for core household types such as unipersonal, nuclear, and single-parent households. While full independence for external validation is limited by data

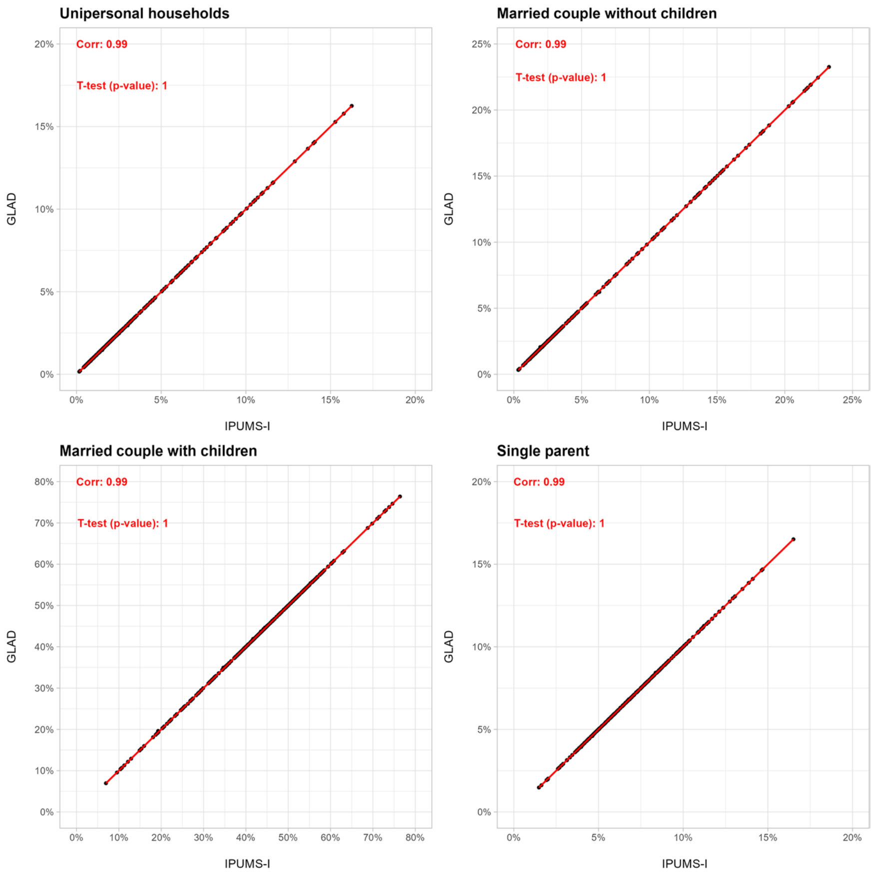

Fig. 6 Validation of the GLAD.

constraints, the strong correlation and non-significant differences in means provide reasonable assurance that our approach yields reliable results for these foundational categories. 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 GLAD and the ones from the IPUMS-I is highly linear, suggesting a very good fit of our computations (Fig. 6). 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

Atlas of living arrangements. The second approach we use to validate this new database is not a validation *per se* but rather an exercise in transparency. GLAD users can access the Atlas of Living Arrangements, which is provided alongside the data. This atlas offers a concise overview of the age and sex structure of the microdata underlying GLAD and the results of applying the algorithm to reconstruct living arrangements from an individual perspective. It includes a leaflet with three descriptive visualizations for each sample in the available datasets:

1. A population pyramid with the distribution, in relative terms, of the population included in each sample by living arrangement types, sex and age (either single or groups).
2. A faceted area plot showing the relative distribution of living arrangements by sex and age (either single or groups) without any data smoothing.
3. A line/area plot displaying the proportion of the population living with different types of kin.

Figure 7 corresponds to the leaflet of the sample of the United States 2020 included in the single ages' dataset. The full Atlas can be accessed here: http://bit.ly/4iqY6ol.

The source code of the Atlas is available in the project's GitHub repository (see section **Code Availability**).

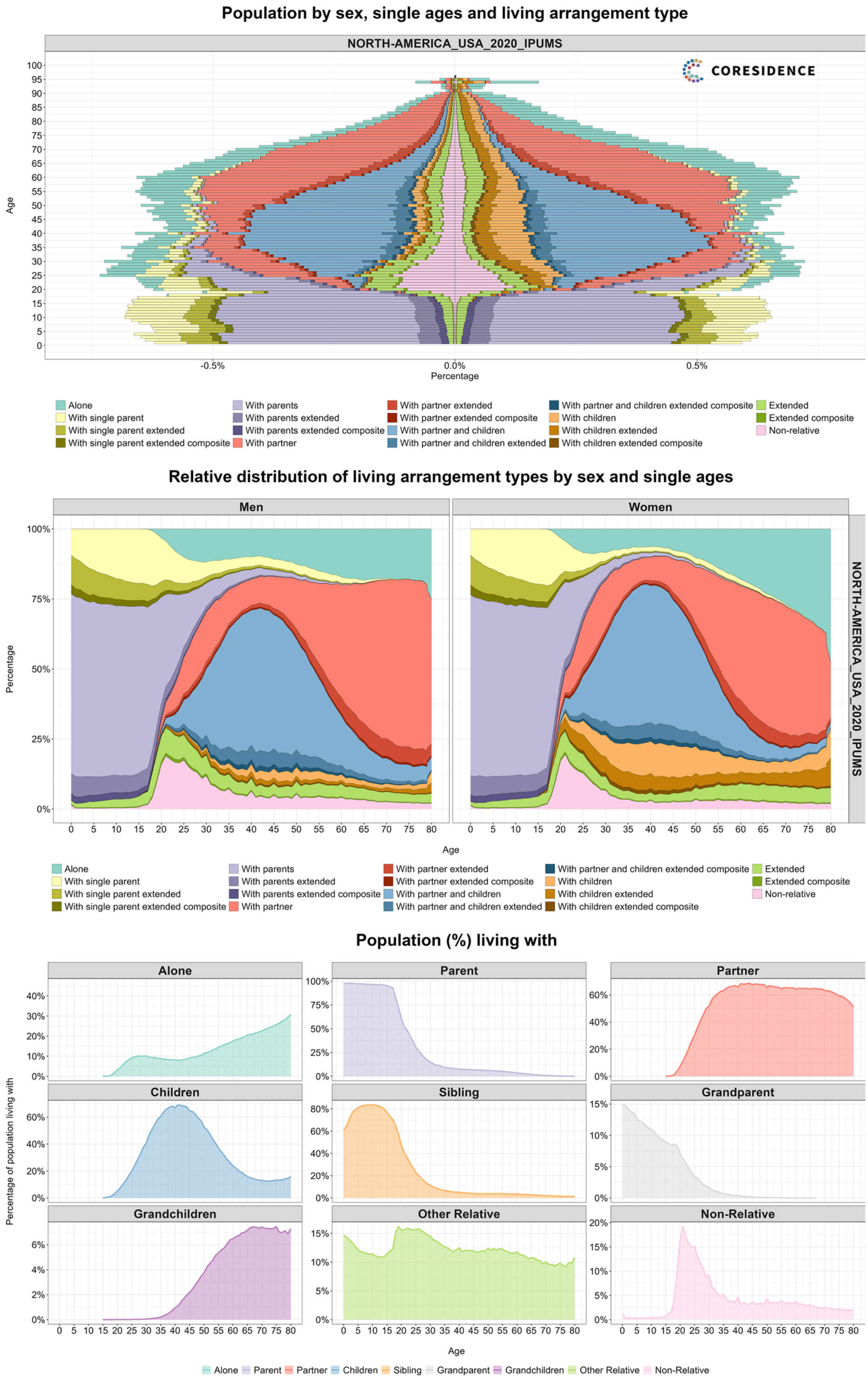

Fig. 7 Atlas of living arrangements, United States 2020.

Code availability

The processing steps to build the three datasets composing the GLAD were carried out in R, utilizing the libraries tidyerse[22], haven[23], labelled[24], tibble[25] and doparallel[26]. All the code is available on the GitHub repository of this project: https://github.com/JuanGaleano/CORESIDENCE.

Received: 21 March 2025; Accepted: 4 August 2025;
Published online: 26 August 2025

References

1. Furstenberg, F. F. *Destinies of the Disadvantaged: The Politics of Teen Childbearing*. https://www.jstor.org/stable/10.7758/9781610442343 Russell Sage Foundation (2007).
2. Elder, G. H. Jr. The Life Course as Developmental Theory. *Child Development* **69**(1), 1–12, https://doi.org/10.2307/1132065 (1998).
3. Furstenberg, F. F. Banking on families: How families generate and distribute social capital. *Journal of Marriage and Family* **67**(4), 809–821, https://doi.org/10.1111/j.1741-3737.2005.00177.x (2005).
4. Bourdieu, P. The Forms of Capital. In Richardson, J. G. (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood Press (1986).
5. Giddens, A. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. http://www.sup.org/books/title/?id=2660 Stanford University Press. (1991).
6. Esping-Andersen, G. Social foundations of postindustrial economies. https://doi.org/10.1093/0198742002.001.0001 Oxford University Press (1999).
7. Esteve, A. *et al*. A global perspective on household size and composition, 1970–2020. *Genus* **80**, 2, https://doi.org/10.1186/s41118-024-00211-6 (2024).
8. Esteve. A. & Reher, D. Trends in Living Arrangements Around the World. *Population and Development Review*, **5**(1) https://onlinelibrary.wiley.com/doi/full/10.1111/padr.12603 (2024).
9. Lesthaeghe, R. The unfolding story of the second demographic transition. *Population and Development Review* **36**(2), 211–251, https://doi.org/10.1111/j.1728-4457.2010.00328.x (2010).
10. Bongaarts, J. & Zimmer, Z. Living Arrangements of Older Adults in the Developing World: An Analysis of Demographic and Health Survey Household Surveys. *Journal of Gerontology: Social Sciences* **57**(3), S145–S157, https://doi.org/10.1093/geronb/57.3.S145 (2002).
11. Bongaarts, J. Fertility Transitions in Developing Countries: Progress or Stagnation. *Studies in Family Planning* **39**(2), 105–110, https://doi.org/10.1111/j.1728-4465.2008.00157.x (2008).
12. United Nations, Department of Economic and Social Affairs, Population Division. (n.d.). *Household size and composition*. United Nations. Retrieved [01/01/2023], from https://www.un.org/development/desa/pd/data/household-size-and-composition.
13. Galeano, J. *et al*. CORESIDENCE: National and subnational data on household size and composition around the world, 1964–2021. *Sci Data* **11**, 145, https://doi.org/10.1038/s41597-024-02964-3 (2024).
14. Hofferth, S. L. & Casper Lynne, M. *Handbook of Measurement Issues in Family Research*. https://doi.org/10.4324/9780203759622 Mahwah, NJ: Lawrence Erlbaum Associates. (2007).
15. Bignami-Van Assche, S., Boulet, V. & Simard, C.-O. A New Methodological Approach to Study Household Structure From Census and Survey Data. *Sociological Methods & Research* **52**(2), 587–605, https://doi.org/10.1177/0049124120986192 (2023).
16. 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).
17. Minnesota Population Center. IPUMS International [multiple datasets]. Minneapolis, MN: University of Minnesota. https://international.ipums.org/international/ (2023).
18. MacDonald, A. L. IPUMS International: A review and future prospects of a unique global statistical cooperation programme. *Statistical Journal of the IAOS* **32**(4), 715–727, https://journals.sagepub.com/doi/abs/10.3233/SJI-161022 (2016).
19. Eurostat. European Union Labour Force Survey [multiple datasets]. European Union (2023).
20. Sobek, M., Kennedy, S. *The development of family interrelationship variables for international census data*. Minneapolis, MN: University of Minnesota, https://assets.ipums.org/_files/mpc/wp2009-02.pdf (2009).
21. Galeano, J. & Esteve, A. CORESIDENCE_GLAD: The Global Living Arrangements Database, 1960-2021 (Versión V1) [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.15038210 (2025).
22. Wickham, H. *et al*. Welcome to the tidyverse. *Journal of Open Source Software* **4**(43), 1686, https://doi.org/10.21105/joss.01686 (2019).
23. 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).
24. Christoph, J. & Lazarevic, M. *labelled: Manipulating Labelled Data*. R package version 2.8.0. https://CRAN.R-project.org/package=labelled (2020).
25. Müller, K. & Wickham, H. tibble: *Simple Data Frames*. R package version 3.1.4. https://CRAN.R-project.org/package=tibble (2021).
26. Corporation M, Weston S. _doParallel: Foreach Parallel Adaptor for the 'parallel' Package_. R. package version 1.0.17, https://CRAN.R-project.org/package=doParallel (2022).

Acknowledgements

We thank Guillem Esteve for his initial proposal on how to compute the living arrangement variable, which was implemented in the creation of this database. We are also grateful to Paolo Marangio and the Barcelona Supercomputing Centre for their support in processing the microdata. Finally, we extend our thanks to the members of the CORESIDENCE team for their valuable comments and suggestions throughout the development of GLAD. The CORESIDENCE project is funded by the European Research Council (ERC). Advanced Grant. Reference: HE-ERC-2021-AdG-GA No 101052787-CORESIDENCE. The project leading to these results has also received funding from "la Caixa" Foundation under the project code LCF/PR/OB23/90000002.

Author contributions

Albert Esteve is the principal investigator of the CORESIDENCE team, he conceived the project, co-designed the analytic strategy and co-authored the initial manuscript. Juan Galeano, co-designed the analytic strategy, developed the algorithm for reconstructing living arrangements from an individual-based perspective, processed the data, wrote the R code for building GLAD, co-authored the initial manuscript and prepared the figures included in this article and in the Atlas of Living Arrangements.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.