



DATA MANAGEMENT PLAN

JANUARY 2022



Prepared by: Andra-Lucia Martinescu
Co-Founder, Research Coordinator



02

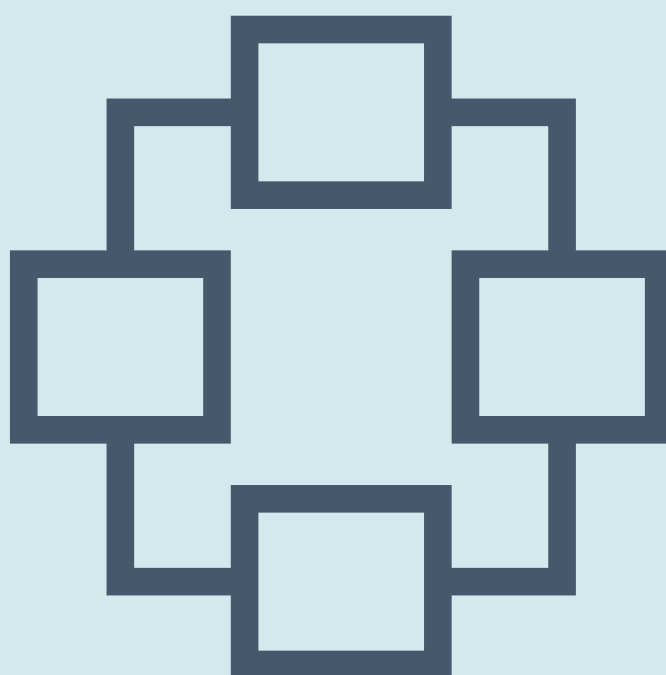
THE DIASPORA INITIATIVE

TDI is committed to high-standard, rigorous, and innovative research. Our approach to mapping diasporas involves a mix of quantitative and qualitative methods, as well as a computational dimension, which enables user interactivity for network visualisation.

THE DIASPORA ECO-SYSTEM

NETWORKS

IMPACT

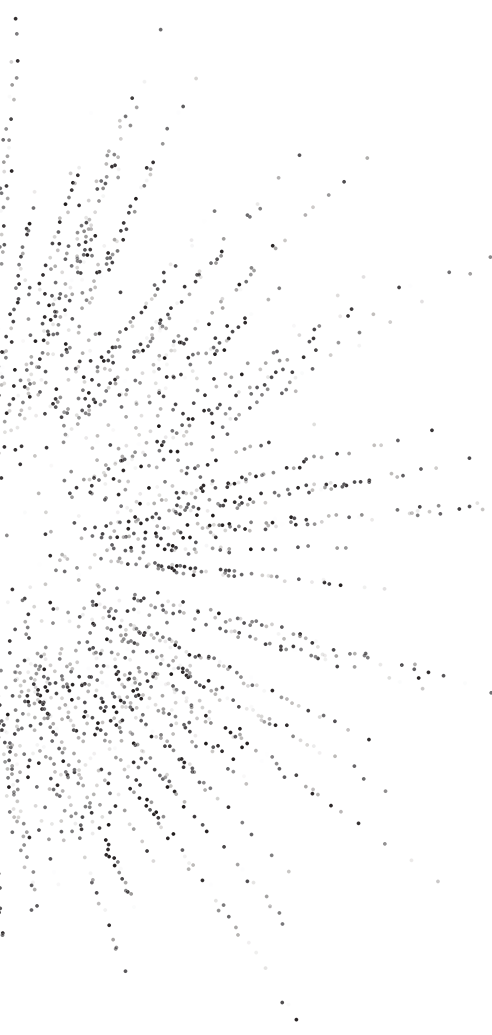


GOVERNANCE

DEMOGRAPHY

03

MATRIX



How many nationals live, work, and study abroad, at any given time? How do they associate and why? How do (diaspora) entities cooperate? With what scope, and to what extent? Can we measure the impact of cooperation? If so, how? What are the rapports with institutions / public administrations (local/national/regional), in both home and host countries? What is the level of participation in home and/or host societies' public spheres?

What physical and affective geographies do diasporas inhabit?

Although not exhaustive in reproducing a diaspora's profile, this framework ([matrix above](#)) enables a context-sensitive approach to thoroughly mapping the presence and impact of diaspora communities, as diverse as they may be.

A. APPROACH



A.1. PRIMARY DATA COLLECTION & INTERPRETATION

A.1.1. DIASPORA NETWORKS & ASSOCIATIVE PATTERNS (THE RELATIONAL DATASETS)



When we decided to map the Romanian diaspora, data was scarce, if not entirely absent. One of our main objectives is to rigorously collect and systematise primary data for interpretation, thereby addressing existing gaps and ensuring replicability. To mediate an understanding of how Romanians associate or organise themselves abroad, forming community associations, charities, volunteer networks and hubs, online groups and platforms, Romanian schools, and Orthodox parishes etc. we created the first (and perhaps only) relational database codifying geographically, by scope of activity, and type all diaspora entities whether formally registered associations or informal groups, both online and offline, per country of residence. Cooperation is assessed based on the frequency and strength (conferred by the weight) of interactions between different actors (or entities). We chose not to limit the scope to only mapping cooperation between diaspora organisations, and as such the dataset includes institutions, civil society, media and the private sector from both home and host countries.

05

This dimension enables a rigorous and more profound appraisal of diaspora participation in the public sphere, of the rapports and interactions with institutions and public authorities, as well as the degree of political representation. The premises for using diaspora organisations and groups (aggregate entities) instead of person-based networks are manifold:

§ Diaspora associations (formal or informal, online and/or offline) provide an interface between the communities they represent (at large) and the public, institutional sphere, in both home and host countries. They can (partially) indicate the levels of community cohesiveness and voluntary participation. The leadership and membership composition may change over the years, but organisations tend to survive / have longer life cycles.

§ Based on the scope of their activity or interest, we can infer which areas of cooperation are prominent and tend to attract most participation (interest), with shifts noticeable over time. For instance, most diaspora organisations cater for the cultural component, that of preserving and promoting the ethno-cultural and linguistic identity of those abroad, organising thematic events (promoting folklore, cuisine, and traditions). Other organisations have a more prominent civic component, launching information campaigns, petitioning and/or liaising with institutions/public authorities as a means of influencing policy agendas.

·§ Diaspora associations perform multiple functions which can also be inferred by analysing the frequency, geographical scope (local, national, bilateral, transnational) and the type of cooperation they engage in (organising / participating at events, initiating civic campaigns, forming local chapters, fundraising, donating, promoting online/offline other groups' events/initiatives/campaigns etc.).

·§ The topology of cooperation (where interactions are situated, geo-location) is also one of the variables used to measure impact, or how influential (diaspora) entities are in a given context. For instance, most associations, which are volunteer based cater for the local community in a city/town, region, or administrative area and therefore have a limited geographical scope. Others, better staffed and with access to diverse sources of funding (which also attest to the organisational capacity) may cooperate with diaspora organisations from other countries, indicating more effective coordination and a higher transnational impact, depending on the agendas they pursue (i.e.: policy change, long-term outlook).

06

In order to identify diaspora entities (representative and with online visibility), we employ triangulation (from multiple sources), and a snow-balling technique, using crawlers to index organisations, groups by online presence (publicly available information accessed through websites, social media community pages, public charity and company registries). The process of codifying the data is followed by mapping the associative environment using a Social Network Analysis software which, upon processing, enables an interactive exploration by node (actor) categories and types of interaction. As for the former, the main category groups have been refined over the years to capture the diaspora eco-system at a granular level (Table 1, page 10).

For each entity (node), we mapped connections (edges) with other actors, partitioning the network based on the scope and types of interaction. We applied queries (filters) to the global network to discern patterns in cooperation, observing for instance, how diaspora associations interact with institutions, whether connections with the public sector in Romania or the UK are more frequent etc. Apart from frequency, the strength of relationships (weight) is also conferred by the scope, location, and type of interaction, which can partly indicate the level of coordination between two or more entities. Online promotion activities (website, social media) carry a lower weight than participating at or organising an event, the latter implying higher degrees of coordination (resources, physical presence). Moreover, factoring in location and scope, the same type of interaction/relationship can take different weight-values. For example, associations from different countries organising an event, implies cooperation at a transnational level, hence higher impact. Whereas associations from the same country organising an event is an instantiation with lower weight.

Receiving financial support, grants, sponsorships (in fact, any material contribution) from another entity and engaging in fundraising initiatives (charity) translate into stronger, more pronounced connections since they tend to indicate a certain level of organisation, outreach, and higher impact. The same applies to other types of relationships: organising/initiating campaigns, charity/civic spin off initiatives, subsidiary/branch of the same entity. If over time repeated interactions between the same entities occurred, the type of connection with the higher-weight value was recorded.

07

The raw data for generating the networks is quantitative. However, for all our analyses and general output, we complement the quantitative dimension with a qualitative component, primary data collected through semi-structured interviews, focus groups and embedded participation (taking part in initiatives, observing cooperation first-hand, participating at community meetings etc., with prior acceptance/invitation). The qualitative data confers extra layers of depth by shedding light on perceptions of belonging, and on inherent cultural and practical norms of cooperation.

Prior to our endeavour, diaspora research has been mainly conducted using a qualitative and survey-based methodological approach, with population statistics (demography) providing the quantitative layer. Computationally, we can now explore a diaspora eco-system dynamically, through multiple prisms. To render network interactivity, we used a Sigma JS plug in, developed by the Oxford Internet Institute (University of Oxford).

A.1.2. DEMOGRAPHIC DATA (MAPPING DEMOGRAPHIC PRESENCE)

Estimating demographic presence is paramount in the mapping of diasporas. Gauging the population size by countries of residence also helps us understand the broader associational activity of diasporic communities, as well as the spatial distribution of diaspora organisations and groups (if they have physical headquarters or traceable activities on the ground).

One metric we employ is the associative density, or the average number of diaspora entities (online and/or offline registered organisations, groups, parishes, weekend schools in the native language, etc.) relative to the population's size, which is indicative of a community's social capital and propensity to participate in volunteer-based activities, forming organisations, attracting membership, or engaging in advocacy. Conversely, a higher density of diaspora associations in certain regions or geographical areas may indicate a growth in the population size, or contextual trends in migration flows.

08

Our approach to interpreting (diaspora) demography takes a critical stance. More specifically we inquire how statistical reporting, and the demographic appraisal of diasporas and migrant communities has evolved over time, to become more nuanced and granular or to include new metrics that reduce the uncertainty of population estimates (if the case). Moreover, demographic data impacts political representation which renders diasporas into a strategic constituency in both home and host countries. Therefore, interpreting how demographic data is being reported, and to what end (i.e.: tailoring the delivery of public services, for electoral purposes or statistical projections) is a fundamental step in our inquiry, and useful for comparative purposes.

Because our approach is longitudinal (observing communities over time) we sought to shape a better, more informed institutional and statistical understanding of diasporas. For instance, our mapping study of the Romanian diaspora in the UK shows how demographic data has been presented by various institutions, and why the reported population size varied to such extent over the years. Two datasets compiled by the Department for Work and Pensions (DfWP) and more recently, the Home Office have been widely used to deduce the population size of Romanian nationals in the UK. The DfWP uses aggregate registrations for the National Insurance Number (by nationality), whilst the Home Office records the number of applications for the settlement scheme (permanent residence). Although the datasets are useful when applied to specific contexts, the true population size cannot be reliably inferred from neither.

Nevertheless, in 2021 the Office for National Statistics (ONS) conducted a nation-wide Census, a process in which The Diaspora Initiative participated as a community engagement partner. The scope was to encourage wide participation from Romanian resident communities in this otherwise compulsory exercise, as well as to provide feedback on structuring the Census output.

09

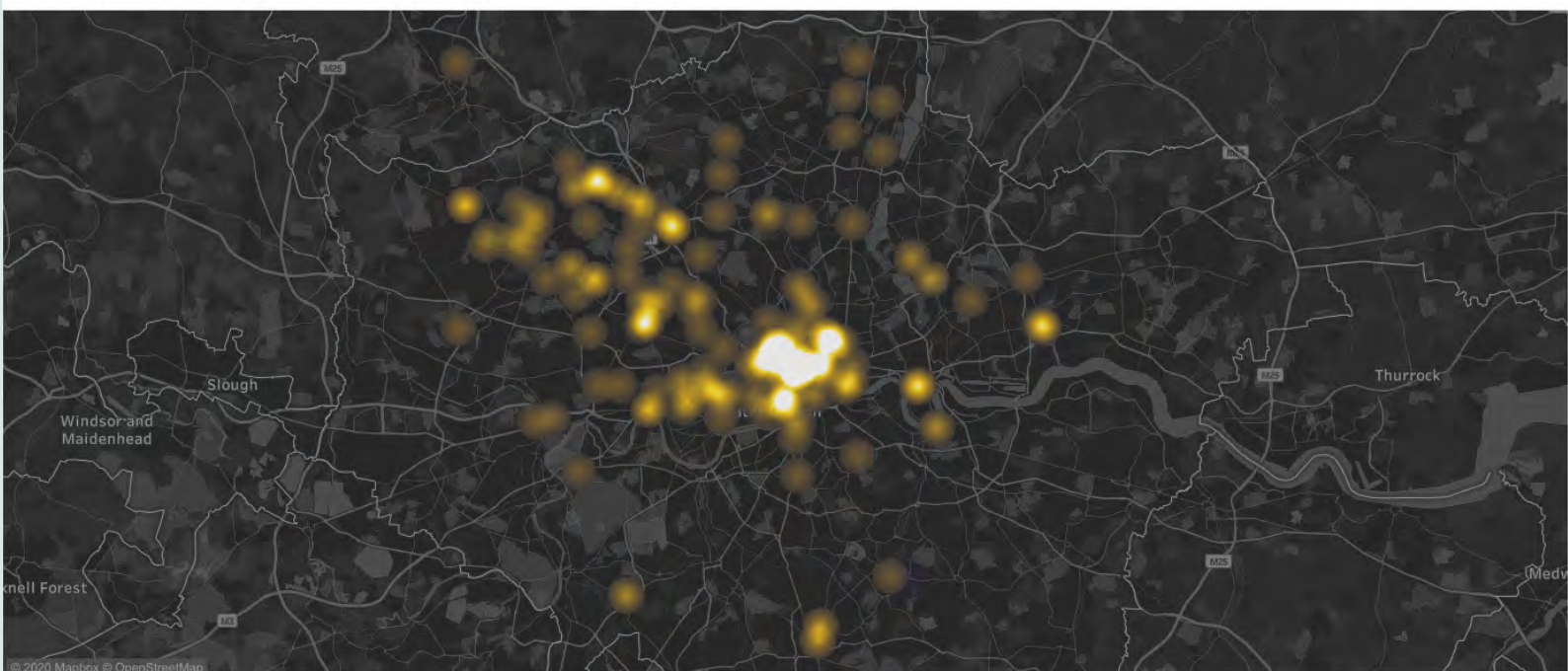
Unlike other diasporas (French, Irish or Italian), the total number of Romanians abroad (emigrants) is extremely difficult to gauge, and depending on the source, the reporting fluctuates between approximately 800,000 and over 5 million. Unfortunately, Romania's national statistics are not reliable, particularly for migration data, due to significant variations in the reporting of outflows (even between different institutional bodies). In effect, the lack of reliable migration statistics affects diaspora representation (in Romania's Parliament those abroad are under-represented), as well as the effectiveness of diaspora-oriented policies and strategies. Thus,

§ For aggregate measures and migration flows we use population estimates by OECD, IOM (the International Organisation for Migration), UN-DESA and EU-Stat.

§ To gauge the migrant stock by country of residence we use national statistics (census data, population surveys and projections etc.).

§ For mapping population and association density the datasets are plotted in geographical analysis software (ArcGis, OpenStreet Maps, Tableau Public).

Greater London - Density of recorded entities



10

CATEGORY	SUB-CATEGORY	TYPE (Examples)
ACADEMIA/ EDUCATION	Academic/educational institutions (country of residence)	Universities, schools, research centres & independent institutes etc.
	Academic/educational institutions (Romania)	
CIVIL SOCIETY	NGO Sector (Romania)	NGOs, local chapters, foundations, charities (national and/or local) etc.
	NGO Sector (country of residence)	
	NGO Sector (Other)	
DIASPORA	Diaspora Entities	Registered diaspora associations, community organisations, Romanian language schools (weekend), student chapters, volunteer-network, online groups & platforms etc.
	Diaspora Entities (Other diasporas)	
MEDIA	Media (RO)	News & media outlets, information portals, investigative collaboratives, professional journalistic associations etc.
	Media (Diaspora Journalism)	
	Religious Sector (Media)	
	Media (country of residence/international)	
PRIVATE SECTOR/ BUSINESS	Businesses	Bilateral trade/professional associations, chambers of commerce, diaspora businesses, etc.
	Business/Professional Associations	
PUBLIC SECTOR/ INSTITUTIONS	Public Sector (Romania)	National ministries and bodies, local authorities and local/regional public administrations, institutions for external representation (embassies, consulates, honorary consulates), etc.
	Public Sector (country of residence/ local administration)	
	Public Sector (country of residence/ national level)	
	Public Sector (Other)	
RELIGION	Religious Sector (Institutions)	Churches and parishes abroad (Orthodox & other denominations), ecumenical associations & charities, weekend schools (hosted by parishes),
	Religious Sector (Diaspora)	
	Religious Sector (other denominations)	
	Religious Sector (Charities)	

Table 1 – Node/actor categories, sub-categories & types (2022 Update)

A.1.3. METRICS FOR IMPACT ASSESSMENT

The table below exemplifies a range of metrics we use to assess the impact of diaspora organisations globally and within specific networks (i.e.: per country of residence).

In the interactive network, upon node selection, a side panel shows all the metrics and values we used to analyse the impact of an entity, such as influence, authority, information brokering, connectedness, and collaborative drive (an organisation’s propensity to cooperate locally, regionally, nationally, and transnationally).



CONCEPT	VALUE	DESCRIPTION
Modularity Class	Between -1 and 1	Employed to detect communities within a network, and cohesion inside the community higher than outside (more edges or interactions forming inside a cluster of entities rather than outside).
		Geographical proximity (organisations, groups in the same city/town, region etc.), similar goals, and areas of interest may drive the formation of communities. Nodes (actors/entities) can be part of more than 1 community.
Degree Centrality	Values are network dependent.	A measure that provides a ranking of the most important nodes (entities) in a network. It is conferred by counting the number of edges (links) a node has. To be used in conjunction with the following measure.
Eigenvector Centrality		This centrality index measures the influence of a node within the network, not solely by the number of connections, but based on the centrality values of that node's links. A high eigenvector score means that a node is linked to other nodes with high centrality scores.
Closeness Centrality (variant > Harmonic Centrality)		It is a measure of how often a node appears on the shortest paths between nodes in the network, by computing the average distance from a given node to all other nodes. With this measure we can estimate how fast the flow of information would be through a given node to other nodes, thereby identifying hubs of information dissemination.
Eccentricity		Represents the maximum distance from a node to any other node in the network, the reciprocal of eccentricity used a measure of importance.
Betweenness Centrality		Detects the influence a node has over the flow of information in a network graph. The algorithm calculates unweighted shortest paths between all pairs of nodes, identifying those nodes that serve as bridges linking parts/sections of the network. It may be indicative of entities that facilitate the flow of information, brokering connections, and generally demonstrating high outreach capacity. Equally, this may be a measure of others' dependence on a given node, and therefore is regarded as a measure of potential control, access efficiency and/or independence from the intermediaries' potential control.
Clustering Coefficient	1 (neighbourhood is fully connected) Close to 0 (hardly any connections in the neighbourhood)	A measure of the degree to which nodes (entities) in a graph tend to cluster together, calculating how connected a vertex's (node) neighbours are to one another.

A.1.4. DATA ON DIASPORA GOVERNANCE

Researching diaspora governance structures comparatively enables us to evaluate how sending states (countries of origin) engage and support diasporas: what institutional infrastructures are in place, what are their mandates and budgets, and how governmental bodies support diaspora initiatives, by overviewing trends and patterns in diaspora funding. One facet of our inquiry focuses on the degree to which diaspora governance acts as an opportunity structure for those abroad: responding to their needs and aspirations, strengthening the capacity of organisations, funding diaspora projects and initiatives and finally, upholding good governance norms.

For the comparative dimension of diaspora governance, we consult government websites to extract information which we then systematise in a descriptive database: diaspora institutions, whether governance is centralised or devolved, the level of inter-institutional coordination, diaspora strategies, and areas of interest. In essence, we systematically review a government's priorities in relation to its diaspora, how these priorities are translated into policies or diaspora strategies and with what outcomes. Using this approach we analysed diaspora governance in Italy, France, Ireland, and the UK, countries with vast diaspora infrastructures and impactful engagement strategies.

We also produced the first and most comprehensive dataset on diaspora governance in Romania, analysing the distribution of diaspora funding, budgetary allocations, reviewing institutional performance (the Department for Romanians Abroad, Departamentul pentru Români de Pretutindeni, DRP), and structural changes over the years, more specifically how the institutional mandate changed over time and with what consequences. The initial scope of our inquiry expanded once we became familiarised with the funding instrument, and how it is actually reported to the public. We developed a monitoring mechanism tracking which diaspora projects have been funded between 2014 and 2021 (to be updated annually), who were the recipients (registered diaspora organisations/ persons by country of residence), what sums were initially approved and subsequently reimbursed (upon project completion).

The type of information we needed was not publicly available or lacked updates and standardisation. To this end, we filed multiple Freedom of Information Requests for the data to be released (based on L544/2001). For funding recipients, only the projects' initially approved sum was reported on the institution's website, not the final reimbursement, which inflates the actual financial support afforded to diaspora organisations. Since the data we received was incomplete and not standardised for statistical analysis (mostly scanned copies that could not be supported in pdf editors) we created a database that centralises all diaspora projects, both approved and reimbursed. An overview of the quantitative and qualitative data is rendered below:

14

Data	Scope	Sources	Details
Quantitative	<ul style="list-style-type: none"> § Institutional Budgets (aggregated and detailed) § Distribution of budgetary expenses (budget allocation) per categories and subcategories. 	Ministry of Public Finance	Accessed through the transparency online portal. In pdf. format, using institutional classification (codes which required deciphering). The detailed budget for 2020 could not be accessed.
Quantitative	<ul style="list-style-type: none"> § Funding allocated for diaspora projects § Recipients of funding by country of residence § Sums requested and allocated initially upon project approval. § Sums reimbursed upon project completion. 	<ul style="list-style-type: none"> § Department for Romanians Abroad (DRP) § Ministry of Foreign Affairs § Department for Relations with the Republic of Moldova § Government General Secretariate 	DRP has changed its structure and mandate over the years, so it becomes difficult tracing public and diaspora funding particularly if the information is not centralised or digitally stored. Obtained through L544 petitions for access to information of public interest.
Qualitative	<ul style="list-style-type: none"> § Overview of diaspora strategies & policies, areas of priority for funding. 	<ul style="list-style-type: none"> § Department for Romanians Abroad § Department for Relations with the Republic of Moldova § Other government websites. 	
Qualitative	<ul style="list-style-type: none"> § Assessing the overall experience of recipients with the funding application and funding in general. 	<ul style="list-style-type: none"> § Semi-structured interviews and focus groups with funding recipients 	If the semi-structured interviews or focus groups are conducted online, we use the secure <u>Webex</u> platform, meetings recorded upon prior approval for transcribing purposes.

We intend to update and publicise the data yearly to capture changes over time. The datasets we compiled indicate the date of collection/update, as well as a descriptive overview of missing or incomplete data. In the visualisation tool (dashboard created in Tableau Public for user interactivity) we also included explanations on financial/budgetary classifications, missing data, years covered by the analysis and sources, to facilitate public/research use.

15

A.2. DATA QUALITY, FORMATS, STANDARD DOCUMENTATION & UPDATES

A.2.1. DATA FORMATS & CONTEXTUAL DETAILS

For qualitative data, editing and text-based documents we use Microsoft Word and `.txt`, `.docx` and `.doc` extensions. The semi-structured interview transcripts are password protected using be-spoke encryption software to safeguard the anonymity of participants. Quantitative data and outputs are saved in `.xlsx`, `.csv`, `.sav`, `.tdsx`, `.twbx` file formats. The relational datasets (for the Social Network Analysis) are saved and stored in `.gephix` and `.js` formats. For audio files we use `MP3`, `MP4` and `.wav` formats, and Quicktime Movie for video files. These are also supported by the Discourse Analysis processing tool, NVIVO.

All the files and datasets have a standard documentation to facilitate processing and public understanding. The Tableau visualisations contain detailed explanations about the data collection, interpretation, and use, to enhance and facilitate user interactivity.

A.2.2. DATA UPDATES

Our longitudinal approach to researching diasporas means that our databases will be regularly updated, on a yearly basis to capture changes in the networks, following established data collection protocols across the matrix: for demographic presence, associative patterns and trends (diaspora network analysis), impact assessment & governance (diaspora funding).

B. QUALITY ASSURANCE & BACK-UP PROCEDURES

B.1. SECURITY/STORAGE

Electronic data is stored on a cloud using identifiers and folder codes to locate the files, depending on the project. The Active Directory structure enables file sharing depending on access permissions, classified as either Read-Only or Read/Edit. The interview and focus group transcripts, as well as the proprietary relational datasets have an extra layer of encryption. External users, if the case, will be provided with an access key and assigned to either Directory group.

Segments of the electronic data are also stored and archived on portable electronic devices protected by encryption software. Sensitive data that needs to be transmitted electronically will first be encrypted.

B.2. DATA USAGE PROTOCOLS

Our research can be explored and accessed via our platform, any subsequent data sharing ought to abide by attribution and academic citation standards. Any privately shared information, research related or otherwise will be subject to a Confidential Disclosure Agreement. A Data Use Agreement applies to the downloading, usage and interpretation of proprietary datasets or segments of the datasets. An ethical framework for usage and exploration is in place to avoid risks associated with how the data is used and manipulated or confidentiality breaches (preventing data usage for political purposes, for instance, or access by unauthorised third parties).

C. COPYRIGHT PROTECTION

The databases are original creations and by reason of the original selection and arrangement of contents the databases are protected by copyright law and hence, are **proprietary to The Diaspora Initiative**. Licensing agreements may be in place for further usage and interpretation.

