SIPHER Glossary

Offering clarification of our terminology and use of acronyms.

Term	Explanation
Attribute-rich	Attribute-rich means that for whatever unit of observation (e.g. an individual) there are many variables relating to this unit. For an individual in an attribute-rich dataset there would many variables capturing different pieces of information about this individual. This information can, for example, relate to income, education, employment, health etc.
Causal Loop Diagram	A causal loop diagram is a diagram which consists of variables (also known as nodes) and links (also known as arcs). Variables are connected by links, shown as arrows, which capture the direction of the relationship. Next to the arrow is either a '+' or '-' sign which symbolises the type of relationship. Two variables connected by an arrow is taken to mean a change in the variable at the tail of the arrow leads to a change in the variable are the head of the arrow. If an arrow has a '+' sign next to it, it means an increase (decrease) in the first variable leads to an increase (decrease) in the second variable. If an arrow has a '-' sign, it means an increase (decrease) in the first variables leads to a decrease (increase) in the second variable.
Cross-sectional	A cross-sectional measurement or analysis refers to a perspective in which there is only one single measurement. This measurement can represent one point in time, such as one day, one week, one year or one 3-year period. Analysing two or more cross-sectional measurement results in a longitudinal perspective.
Data Zone (DZ)	Data zones (DZ) are census-based, geographical areas in Scotland. Data zones contain approximately the same number of people who live in the area (around 700), but the geographical size can vary. The geographical size of DZs is small in densely populated areas (like a city) and have a larger geographical size in less densely populated areas (like rural areas).
Discrete Choice Experiment	Discrete choice experiments are a preference elicitation technique. They seek to understand and allow for quantification of how individuals trade-off alternatives. Typically, participants are asked to choose between a discrete number of alternatives. Alternatives can

Term	Explanation
	contain a set of associated characteristics and the choice between which alternative is better than the other depends on the underlying preferences of the participants (the weights the individual places on the individual characteristics).
DST	Decision Support Tool – for more visit SIPHER DST
Dynamical Systems Model	This is a mathematical framework used to describe and analyse the behaviour of complex systems over time. It represents how variables in the system change and interact with each other, capturing the dynamics and dependencies within the system. This type of model is widely applied in various fields, including physics, biology, economics, and engineering, to understand and predict the behaviour of real-world systems.
Electoral Ward	Refers to a level of geographical resolution as well as an electoral district in the UK. In total Scotland, England, and Wales are made up of 8,082 wards as of 2022. Information on boundaries, and boundary changes over time are provided as well as geographical codes for maps are provided here: https://geoportal.statistics.gov.uk/.
Equivalent Income	Equivalent income: given a current situation captured in terms of income (y) and non-income aspects (X), equivalent income (y') is the amount of income that, if combined with the best levels of non-income aspects (X*), is as good as the current situation (y, X). It is a preference-based single index of wellbeing in monetary terms that penalises the level of income to reflect any deficiencies in the non-income aspects, based on how important each deficit is.
EQ-5D-5L	**EuroQual 5 Dimensions (Health and Wellbeing Questionnaire)**
EQ-HWB	**EuroQual Health and Wellbeing instrument (Health and Wellbeing Questionnaire)**
General License / Special License	Data provided by UK data service, such as the Understanding Society Survey, often comes with different data access agreements. This is due to the fact that data may differ with respect to amount and character of potentially confidential information - for example with respect to details surrounding an individual's place of residence. Most of the time, the more confidential a data set is - the higher the

	thresholds for accessing it. A general license agreement represents a situation in which data can often be accessed without further restrictions, after agreement to the respective terms and conditions. In contrast, a special license agreement refers to a situation in which data can only ever be accessed after prior approval by the data provider, which often involves sharing a research proposal including a clear justification for why certain safeguarded information is required. Further information to different license types can be found here: https://ukdataservice.ac.uk/help/access-policy/what-data-can-i-get/
Health Inequalities	Health inequalities are the differences in health of different groups within a population. These inequalities are considered unjust, unfair, avoidable and systematic.
HUI	**Health Utility Index**
ICECAP-A	**Icepop Capability measure for Adults (Health and Wellbeing Questionnaire)**
Imputed Data	Data often contains missing values, which can be a challenge for some types of statistical modelling. When a missing data point is replaced with a particular value, for example the overall mean or median, this new value represents an imputed data point. There are various methods to decide the best replacement value, all of which have their strengths and limitations. There are different approaches to imputing missing data. A summary of one of the most common and robust approaches is provided here: https://gking.harvard.edu/amelia
Inclusive Economy (IE)	Improvements across a large number of domains, for example with respect to population health, have historically been attributed to factors surrounding economic growth. Within the past decade, limits to economic growth, for example given by environmental and planetary boundaries, have become increasingly obvious. The idea of an inclusive economy is particularly concerned with economic inclusion, rather than measures of size or growth of an economy. While there are different definitions of the inclusive economy, a number of common attributes have previously been pointed out. Through an iterative, consultative process, SIPHER has selected 13 relevant dimensions and compiled two indicator datasets reflecting these dimensions for local authorities and electoral wards in Scotland, England, and Wales. Further information about SIPHER's approach to defining and developing indicators of an inclusive economy can be

	found in the Technical Report, available here: https://sipher.ac.uk/wp-content/uploads/2022/10/SIPHER-Inclusive-Economy-Indicator-set.pdf. A review outlining different approaches to defining inclusive economies can be found here: https://jech.bmj.com/content/75/11/1129.abstract
Inclusive Growth (IG)	This refers to the broad idea that growth and prosperity should create opportunities for the wider population and that the benefits of growth should be distributed fairly and reduce inequalities. The RSA Commission, for example, defined inclusive growth as 'broad-based growth that enables the widest range of people and places to contribute to economic success, and to benefit from it too' (see page 7, RSA (2016) https://www.thersa.org/reports/emerging-findings-of-the-inclusive-growth-commission, September 2016). However, inclusive growth remains a fuzzy concept, with some versions emphasising the need for inclusion within existing growth models (a 'growth plus' version), while others emphasise the need to change the economy so that poverty and inequalities are reduced by design (see the discussion starting page 6 from the report Achieving Inclusive Growth in Greater Manchester: what can be done? http://hummedia.manchester.ac.uk/institutes/mui/igau/IGAU-Consultation-Report.pdf)
Inequality Aversion	The extent to which an equal distribution of an outcome (such as income or health) is preferred over an unequal distribution. Where people are averse to inequality, a less-but-equal distribution is preferred over a more-but-unequal distribution.
Local Authority (LA)	Refers to a level of geographical resolution as well as a level of local government funding in the UK. In total Scotland, England, and Wales are made up of 363 local authorities (district level / lower tier). Information on boundaries, and boundary changes over time are provided as well as geographical codes for maps are provided here: https://geoportal.statistics.gov.uk/.
Longitudinal	Longitudinal data are data that were collected repeatedly over time from the same subject. This is in contrast to cross-sectional data which collects data once. For example, the Understanding Society survey data are longitudinal data as the same households participate in the survey and answer the same questions annually.

Lower Super Output Area (LSOA)	Lower Super Output Areas (LSOA) are census-based, geographical areas in England and Wales. LSOAs contain approximately the same number of people who live in the area but the geographical size can vary - their geographical size is small in densely populated areas (like a city) and have a larger geographical size in less densely populated areas (like rural areas).
Macro-level modelling	Macro-level modelling produces aggregate outputs and provides information about larger population groups, which are often geographical areas or countries.
Matlab	Matlab is a commercial statistical and mathematical platform which comes with its own syntax. More information on Matlab can be found on the providers website: https://www.mathworks.com/products/matlab.html
Microsimulation	A modelling technique used to simulate the behaviour and interactions of individuals offering an insight into the impact of policy interventions. It requires the creation of a synthetic population data source providing a digital twin of the individuals to be examined with their specific attributes. By simulating their decisions, actions, and interactions over time researchers can explore the effects of different policies, interventions, or scenarios at the individual level and aggregate this to understand community impacts.
Micro-level modelling	Micro-level means the model seeks to tell us something about individuals, groups of individuals or very small areas.
Model output	Model output reflects a situation in which a result was obtained through a statistical process, contrasing situations in which we deal with raw input data. Even descriptive statistics can, in some situations, represent model output. For example, when obtained from synthetic data sources such as the SIPHER Synthetic Population. Model output should always be interpreted and discussed within the context of its creation, acknowledging the strenghts and limitations of the creation process.
MSOA	Middle Layer Super Output Area
NRS	National Records of Scotland

Non-synthetic data	Non-synthetic data represents data that were gathered through various data collection methods - for example as part of a register-based population system or a survey. Non-synthetic data are reflecting observations of true real-world units, such as real individuals or geographical areas.
ONS	**Office for National Statistics**
ONS-4	**Office for National Statistics - 4 (Health and Wellbeing Questionnaire)**
ORDA	**Online Research Data**
Python	Python is a free and publicly available, general-purpose programming language. It is particularly useful when processing large amounts of data. More information on Python can be found here: https://www.python.org/
QALE	**Quality-adjusted life expectancy**
R	R is a free and publicly available programming language. R has a large, interdisciplinary user base and is particularly well-suited for statistical analyses and data visualisation. The R-project: https://www.r-project.org/
Register-based System	A register-based system refers to a situation in which each individual of a country can be traced exactly - for example due to a unique personal identification number - across multiple sources of routinely collected administrative data. These sources could cover for example: education, tax, prescriptions, hospital admissions, fertility, housing, cause of death etc. This system is common in the Nordic countries, were a register-based systems has been in place since the 1960s. Within the past years, a strong move towards data linkage has taken place in Scotland and the UK and with respect to NHS data. Historically, register-based data were collected routinely for administrative purposes, rather than with a particular research purpose in mind. For Scotland, eDRIS provides the infrastructure to many registers - often referred to as national data safe havens: https://www.isdscotland.org/products-and-services/edris/ The system is more decentralised in England and Wales

Shortform 12 (SF-12)	Short Form 12 Health Survey and its mental and physical health scores MCS and PCS Shortform 12 (SF-12) is a standardised questionnaire consisting of 12 questions related to the responding individual's health. From these responses, two summary scores known as the physical summary component (PCS) and mental component summary (MCS) can be calculated for the individual. This differentiation into physical and mental health enables a holistic perspective on an individual's health. This holistic perspective presents an advantage over more traditional health measures which are often only indicative of an individual's risk of dying or the reflection of a broadly captured self-rated health status. There are different version of the SF-12 available. For example, Understanding Society uses the SF12v2. The SF-12 instrument has been developed by the company quality metric: https://www.qualitymetric.com/sf-12v2-pro-health-survey-lp/
Simulation	A simulation refers to a study design in which hypothetical changes of the observed units are modelled. Units under study can be individuals, households, administrative units such as UK Local Authority Districts, or entire countries. Simulations come with the advantage that shocks or interventions can be testes and examined before taking place. In turn, this requires that we have a certain amount of information about the particular event or the policy intervention. Simulations have a particular advantage over traditional study designs as shocks can be studied before they have occurred and interventions can be tested before any money is spent. A major drawback of simulations is the fact that despite having as many insights into potential behaviour as possible, in reality, our units of observation might still not behave the way we have predicted it.
Stata	Stata is a commercial statistical software which comes with its own syntax. More information on stata can be found on the providers website: https://www.stata.com/
Synthetic Data	Synthetic data are artificially generated data, which may or may not draw upon existing data. Synthetic data are created in SIPHER because there is a need for it. In the SIPHER project, synthetic data are generated and processed by WS3 and used in WS4, WS5 and WS7 models.
Systematic Review	A scientific method to identify, structure, condense and summarise previous research findings and other sources of evidence with respect to a particular topic. This approach helps to eliminate

	subjectivity and selectivity with respect to a studied topic. Results of systematic reviews help us to understand the current state of research as well as future directions.
UK Data Service	The UK Data Service is the UK's largest institution collecting and providing economic, population and social research data. The data provided via UK data service is typically used for research, teaching, learning and public benefit. Access to many data collections, including the Understanding Society survey data, is provided via the website of UK data service. While some data is easily available, other data - in particular with respect to sensitive information - might be subject to special license agreements and other forms of safeguarding. More about the UK Data Service can be found here: https://ukdataservice.ac.uk/ An overview of domains covered in UK data service data sets is provided here: https://ukdataservice.ac.uk/help/access-policy/what-data-can-i-get/
Understanding Society (US)	Understanding Society, also known as the UK Household Longitudinal Study, is a longitudinal study which follows participants over a time. Households are selected to join the study, and all members of the household participate in the study. The survey is conducted annually and collects data from participants on their lives on a set of wide-ranging questions. The survey began in 2009 and substituted the British Household Panel Survey (BSPS) - which ran from 1991 to 2009. Harmonised BSPS waves are available and enable an even longer time frame when combined with Understanding Society. More about the study can be found here: https://www.understandingsociety.ac.uk/about/about-the-study Access to the study is provided via UK data Service - a link to the General License Version: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6614#!/details
Variable	The terminology variable can imply different concepts. Typically as well as in this context, the terminology refers to an object which captures any type of information (e.g., name or spending) with respect to an observational unit of interest (e.g., a person, a household, a Local Authority District, the Scottish Government), and with respect to a particular point in time (e.g., yesterday, today, last year).
Ward	See Electoral Ward.

WEMWBS	**Warwick-Edinburgh Mental Wellbeing Scale**
Workstrand (WS)	The SIPHER project is divided into eight workstrands. Each Workstrand has a particular focus on either qualitative and quantitative questions. This diversity represents a particular strength of SIPHER as workstrands contribute to each other in a supportive and iterative manner, for example: theory is informing statistical modelling and results of statistical modelling can change perspective on existing theories. A summary of SIPHER workstrands and their synergistic interaction is described here in more detail: https://wellcomeopenresearch.org/articles/4-174/v1 and available https://www.gla.ac.uk/research/az/sipher/sipherwheel-workstrands/