Chapter 5
From Concept to Tangibility

The techniques have galloped ahead of the concepts. We have moved away from studying the complexity of the organism; from processes and organisation to composition.

[Commenting that growing use of new technologies and techniques, from molecular biology to genomics, has proved a mixed blessing.]

Sir James Black

Quoted in Andrew Jack, "An Acute Talent for Innovation", *Financial Times* (1 February 2009).

As in all studies and processes, for every great work, there are some rules to follow: Newton's Third Law states that: "For every action there is an equal and opposite reaction". In sociological terms, as every citizen in society has his/her own rights and obligations, one is accorded rights to reside in a country but has the obligations to abide by its rules and regulations. These are commonly referred to as laws. Scientific analysis and statistics also have their rules.

Basic Concepts

Let us take an overview of some basic concepts, before embarking on the process. This process entails a flow from the ideal "idea" to the actual, on-the-ground, tangible aspects (Reeve, 1997).

- **From Ideas and Concepts....**
  - Where does one start from? The idea is to form an abstract ideal image in one's mind, based on readings and experiences. This mental image sets the ground through its summarisation of sets of similar observations or ideas. The process this step takes covers a sequential series of events termed conceptualisation, entitation, quantification and measurement. These are discussed later on in the chapter.

- **To Operations...**
  - Getting those ideas a bit closer to reality
  - One needs to identify the value of the cases in a variable

- **To Transformation...**
  - Initiating the process of linking abstract concepts into empirical indicators that can be analysed through hard and measureable steps
  - Measurement issues need to be brought to the fore through the implementation of the rules

- **To Variables**
  - Identify the characteristics or properties of the variables
    - Variables are entities that **vary**
  - Note that each variable produce two or more different values
  - These variables can be categorised as *(now very discernable through the previous chapters)*:
    - Quantitative – measures a quantity or amount
    - Qualitative – measures a quality or category
- **And finally to Indicators**
  - Indicators are self-descriptive. They indicate how to measure a value.
  - These should also indicate the W6H
  - Indicators also describe the operation (question) used to indicate the value of cases in a variable

Two examples that follow the main steps of the above are the following:

Concepts	Variables	Indicators
Drug taking	Frequency of drug use	"How many times in the last year has the offender taken drugs?"
Air Pollution	Vehicles per length of roads	"How many vehicles are registered per kilometre road length"

Let us now review some data issues, particularly the first rule one must keep in mind!

Data issues – The structures

Before we even start to understand how to get from **A** (ideas) to **B** (the variable and subsequently the indicator), it is best to start from the first basic rule…

The Rubbish Rule!

Or better known as…

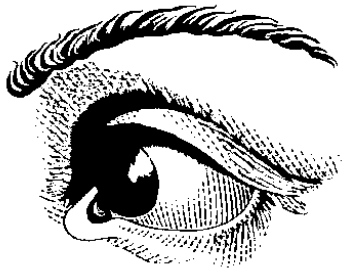
Gallois's Rule

Pierre Gallois described a very realistic rule for computer-based information which can be transposed to any type of data, information or model. The rule states that:

"If you put tomfoolery (rubbish) into a computer, you get tomfoolery (rubbish) out. But the tomfoolery coming out of a very expensive and complicated computer will have attained a sort of respectability and no one will dare criticise it"

THUS:

GARBAGE IN – GARBAGE OUT
(GIGO)



Moral of the story: *Researchers cannot hide behind the technology: but must make an honest evaluation of the validity of the data that is collected.*

As indicated in the previous pages, the process to arrive at a valid variable from a string of ideas is not an easy task. However, taking the process described through database technologies and spatial information systems, one can understand the practical ways described below in order to be able to make sense out of the data created for one's statistical analysis.

To take an abstract to a functional approach, one must start from the concept – the original idea, then structure that idea into something which can be touched in principle, then process that into something tangible on the ground, and finally ensure that it is valid to the item you want to analyse. Once all this is completed, the next step would be to bring all such issues together and enable the modelling function, where the "whole is greater than the sum of the parts".

i. Conceptualisation

Transforming one's idea into a tangible construct requires one to go through a step that allows the researcher to nail down the elusive variable. Think in terms of definition of an idea. If the definition reflects the original idea, then the concept has been established.

This is not always a straightforward task as some social issues are not easily defined. Taking an example from the physical sciences, one can readily identify an orange as having a spherical shape, as being constituted of a series of slices each containing multiple mini-sachets of juicy liquid and that this fruit can be eaten (or drank, if squeezed!).

Now try to define a social construct into a quantifiable and measurable indicator…. Easy? Not really! This is because the boundaries have been defined. These uncertain boundaries are called "fuzzy" concepts – it is not always concretely clear what the definition is. Some examples would include: social segregation, educational need, poverty and areas of special deprivation, among others.

Taking the case of deprivation, one can immediately picture someone suffering from hunger or physical lacking, however… What is deprivation? Is it not a form of poverty? Is there a bias towards one specific type of deprivation? What constitutes deprivation in terms of variables?

The fuzziness here is too real as deprivation means many things and can be analysed using many forms. One has to clarify that deprivation is more complex than poverty, which can be analysed through being short of money (the argument is based on the fact that lack of income results in lack of access to goods and services). Deprivation is based on a model or a group of variables that interlink to help form a tangible variable from the idea.

Take the example of measuring poverty in an apartment block in Valletta. A millionaire might live in an apartment on the tenth storey but is bed-ridden and has no access to a lift. As a result, his social contacts have been severed and he cannot interact with his peers. Though not poor, he is deprived. How would a researcher measure such instances to come up with one definition of deprivation? While building a model is not that easy, some, such as the Index of Deprivation in the UK[1], include hundreds of variables and are very complex. Others may only choose one variable, such as access to the community areas of a town. This process can only be enhanced through review of the literature written on similar circumstances to the ones under study. Then the researcher should review the availability of data for that variable being chosen.

This step requires the rethinking of the process and reworking the choice of ideas until refined enough to elicit the study of a realistic variable.

ii. Entitation

Having taken the mental leap and molded an idea into something that can be measured, the next step in the process is described as entitation. This step describes the way we recognise, understand and define the entities about which we wish to collect data.

What is an entity? An item that can be described and measured in statistical analysis is termed an entity. It spans all the themes under study in research: a single variable to complex models; or single relationships (psychological 1:1 relationships) to complex multiple-organism relationships (sociological multiple-individual interactions).

There are many issues at stake in ensuring the process of entitation. Policy makers may rush to measure things without first pausing to consider whether the same things are worth the trouble. It is essential to ensure that the researcher finds the right things to represent their work with and that they define the boundaries and constitution of the entity as well as the context within which that entity 'lives' (Reeve, 1997).

As an example, one could take spatial development boundaries or the NUTS 5 area boundaries (Refer to Chapter 7 for description of NUTS).

Can anyone tell exactly where Balzan's boundary really is? Where does it end and where does it start? One can equate that to a GPS series of locations on a map, but this might not be so straightforward to someone who has just had the misfortune of being involved in a traffic accident at the traffic lights leading to San Anton Gardens! The boundary is an exact fuzzy location between three NUTS 5 local councils. So, that someone who has just been involved in an accident is perplexed when trying to find out under which police district jurisdiction such an area falls.

[1] http://www.neighbourhood.statistics.gov.uk/dissemination/Info.do?page=analysisandguidance/analysisarticles/indices-of-deprivation.htm

The same issue can be experienced if one wished to organise a barbecue at Mistra Bay. Since such an activity requires a permit, the logical solution is to request a permit from the Mellieha Local Council. Fair enough! However, what happens when one has to state on which exact location such an activity would occur? The left-hand side of the bay (facing the sea) is within the Mellieha area, but the right hand side is within San Pawl il-Bahar's jurisdiction. Since one cannot exactly tell where such an activity would occur (due to various factors such as areas taken up by boat carriages, other barbecues, bathers, etc), and one cannot define exactly where the dividing line is drawn, boundaries becomes practically impossible to define. The most logical solution to the dilemma could have nothing to do with the geographic location of the proposed activity, but would have been taken simply to avoid the long drive up to Mellieha (if driving from the southern localities). The organisers could merely state that the activity would occur in San Pawl il-Bahar territory.

Thus, the simple process of moving from the concept to creating an activity (conceptualisation) to mould that activity (entitation) cannot be held in terms of the exact spatial location but has to be concerned more with the activity itself, the barbecue, that can be measured: Did it occur or not? Did it occur in clearly-discernable Mellieha or clearly-discernable San Pawl il-Bahar area? There is no need to be bogged down with extra detail such as the exact spatial location since that is superfluous to the activity. How many party-goers take a map, a GPS or a SatNav with them to find the exact location of the boundary? Now that would be funny! Especially considering that there is a displacement of significant metres in the technology…

iii. Quantification

The next step takes into account those entities that have been given a form. Can these now be given a measuring scale? Can the values that represent those same entities be measured?

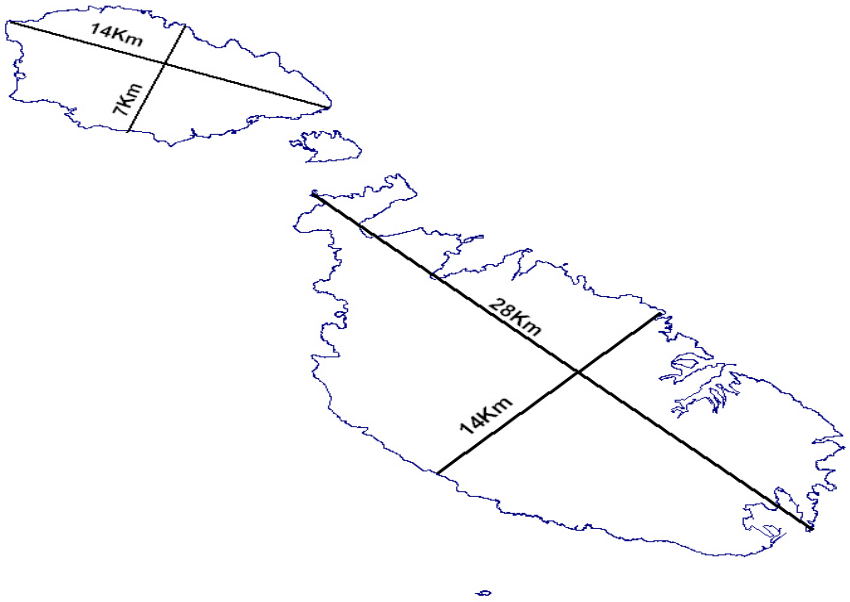
One might say that this is entirely possible.

- How tall is a person? A measuring tape will give the exact length.
- How far is Sicily from Malta? A simple browse through an analogue map or an online browse through a digital map will give you the distance.

But the doubt fiends start creeping in and ensuring that the carpet underneath our feet is not that sturdy any more. The research bugs are attacking one's solidity again. Let us check the following issues based on the above second example.

- Any map will give a distance of 90km from Malta to Sicily, however the researcher has to be more precise: Exactly from where to where has the distance been calculated? From the Valletta Grand Harbour breakwater point to Pozzallo or from any other point to any other point? The exact locations are required. Grab a computer and visit an online map and check this out!
- Another example closer to home concerns that of the geographical construct of the Maltese Islands: a quick review shows that Malta is 28km long and 14km wide, while Gozo is half that – 14km long and 7km wide (Figure 5.1). These figures are easy to remember since they are all divisible by half within the Islands and across the islands – 28/2 equals 14 and 14/2 equals 7! Easy, no? But this just gives one an approximate idea and not an exact measurement since there are no end-point locations given. Quantification requires the exact dimensions pertaining to that measurement.

Figure 5.1: Coastal Distances



Now let us go down research lane. More complex issues are encountered as one moves towards the social domain.

Can these be measured?

- How red is a red car? The eye of the beholder can mislead; red depends on wavelength and perceptual encounters. What is red exactly and how dark or light is dark red or light red? This measurement needs to be clarified before the data is gathered. Whether employing physical, chemical, textual or any other characteristic baseline analysis to ensure that one has a definition of the characteristics of the 'red', the result will hold through as a benchmark for the study and other ones which will be used to compare in the post-study reruns.

Let us take this issue to another level. The same issue discussed in terms of the red factor is encountered in terms of socially-specific data, where data on key variables in the social science research are often missing and one may need a surrogate to measure such.

- Define affluence! What is affluence and how can it be measured? How rich is a rich person? Note that relative poverty may not mean the same thing to two different social groups, so in effect the 'redness' of poverty has to be defined beforehand.

    i. As an example, one can take the case of a farmer who owns a small plot of land in a subsistence-based economy. The farmer can grow his own crop in order to survive. The neighbour, on the other hand has no patch of land and has to beg for his food. The neighbour is definitely poorer in a poor community.
    ii. The second example concerns the case of a yacht owner in Monaco. His yacht is a mere 28 footer and is berthed next to a sleek, gleaming, ultra modern 70 footer. The 28 footer owner is considered a poor fellow compared to his neighbour and could be looked down upon.
    iii. In these two cases the poverty issue is relative to the actors but also across the cases. The 'poor' fellow living in the 28 footer is definitely far richer than the 'rich' farmer working the patch! The latter is utterly poor compared to the rich guys!
    iv. Validity

Once the concept has been transformed into an entity which in turn has been given a measurement structure, then the next step is to ensure that the measurement is valid and that it represents the item under study. The setting up of an indicators list and a lineage process will help the researcher to ensure

that the particular dataset is valid and can be repeatedly gathered, analysed and processed. Indicator lists help the researcher follow the analytical process on a step-by-step basis, which will allow for comparative output analysis.

Ensuring that the data collected is a close representation of the issue is known as an Acid-Test exercise. Follow the rules, and the data processing is considered reliable. However, if your reputation as researcher depends on a dataset of dubious origins bin it!

One item relating to validity regards the use of surrogates.

Surrogate: the substitution of one variable by another corresponding or similar variable.

- Consider the case where a person or a couple would like to conceive but are unable to do so. Cases have been evidenced where a person is contracted to carry the baby for them. The 'carrier' is called a surrogate as she replaces the mother's pregnancy and parturition episodes.
- In social sciences, the use of proxy or surrogate indicators is used in order to help analyse a relationship for which data may not always be available.
- An example would include income data, which question, even in Census takes, has been removed. Such a question may either elicit untruthful or misleading answers. Thus one has to find the best choice that represents that variable, something which is related directly or indirectly, but it must be related. The Census income example would serve as an ideal case through the use of a surrogate variable such as car ownership. The latter may serve as a surrogate for wealth. The car type, cost, number of cars per household, year of production and other related variables serve as a pointer to the level of income.
- Note that any surrogate choice may introduce a measure of doubt as the choice may be deemed subjective by different users.

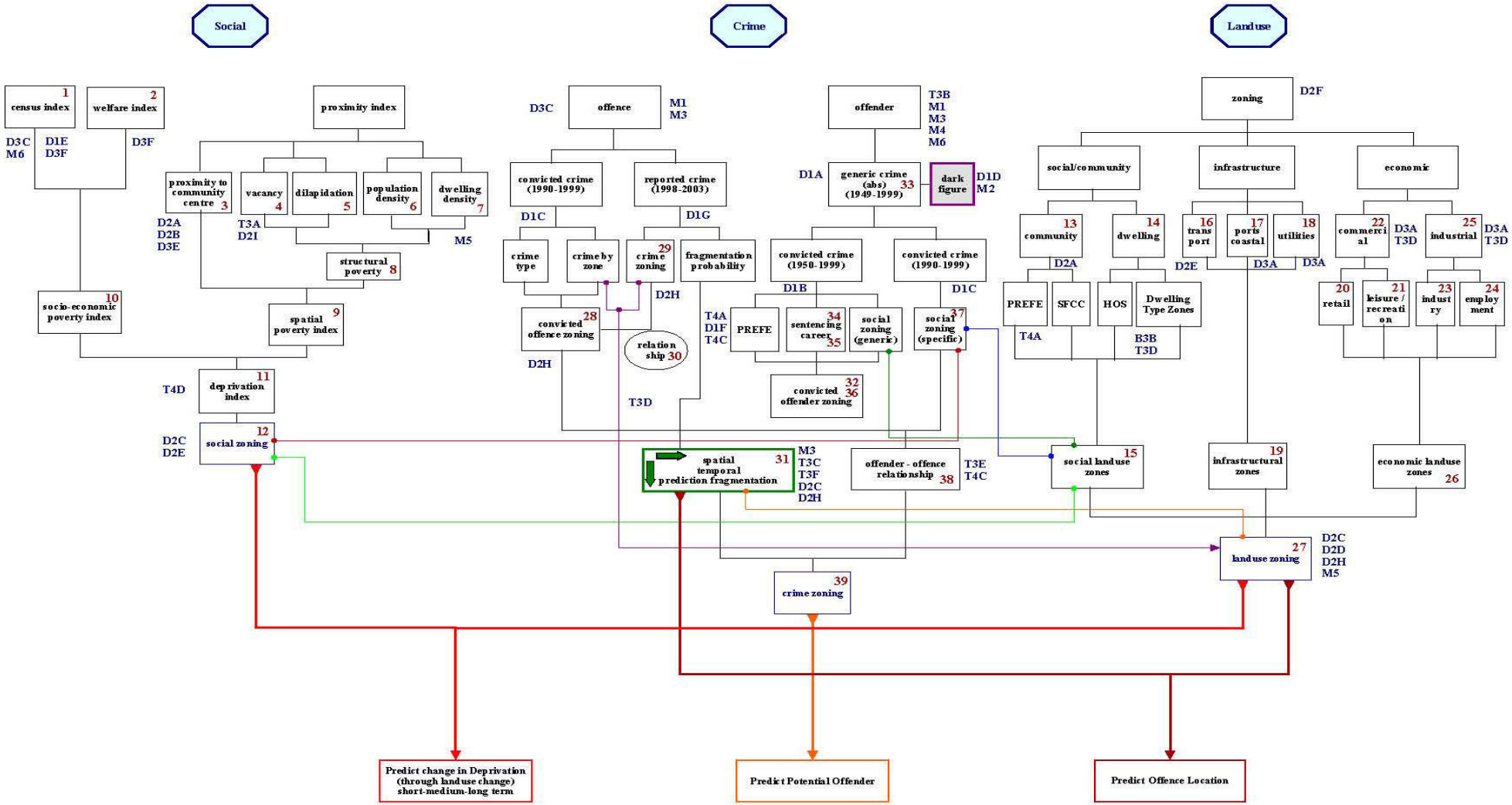
v. Composition

Finally, individual indicators are combined into a composite model and all the data comes together through a process. The composite is a sum of the parts, though in actual fact it is bigger than the sum as information and structures are not emanating from the individual variables but from the linkages between the different variables.

Composition is never an easy process but if approached in the correct way, it easily proves to be a relatively smooth exercise to engage in. This book dedicates a whole chapter (Chapter 8) to the description of how to carry out mind mapping for a research study. The mind map exercise is covered in detail through a model called CRISOLA (Formosa, 2007) which depicts the process described above from conceptualisation to composition (Figure 5.2).

In summary, the composition process allows researchers to develop concepts, visualise variables, create the linkages between the variables and also to identify the statistical measurements required for each linkage, and how they link within themes and across themes.

Figure 5.2: Crisola Model Phase 2 (Formosa, 2007)



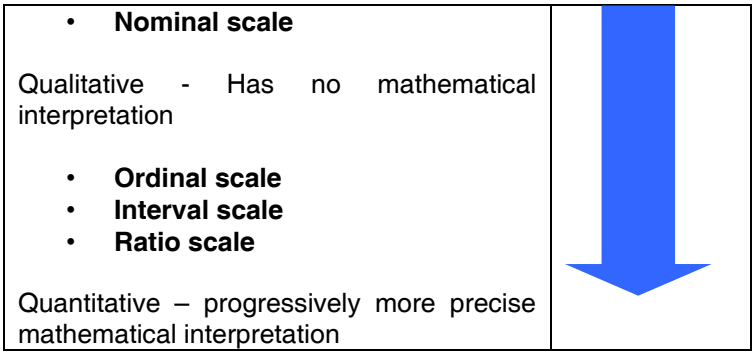
Data Measurement: The scales

Having structured one's thought on the way a research will take place, the researcher is bound to understand which measurement scales to use for each type of query. The previous chapter (Chapter 4) tackled Matrixing and there one could identify the need for the tagging of a measurement scale for every variable.

The Measurement Scales structure follows an ever more detailed and defined process of precision, which is reflected in a mathematical staged process of precision upon which the values of a variable are defined.

There are 4 types of measurement scales, each with the higher level of precision than the previous scale. While Nominal has no mathematical interpretation, Ordinal scale has a higher level of precision, with Interval more precise and Ratio the most precise.

- **Nominal scale**
- **Ordinal scale**
- **Interval scale**
- **Ratio scale**



A description of each would suffice to ensure that one understands which one should be used for the analysis process[2].

Nominal scale

The Nominal scale serves as a pointer to Qualitative variables. There are various characteristics that describe this scale.

- NO quantitative information can be gleaned from this scale
- The scale does not refer to amounts (numbers) but to tags that serve as an identification
- One item in the variable does not signify a greater composition than another item (male is not more important or bigger than female, Attard is not a less important village when compared to Zurrieq)
- The reference to categories is paramount, therefore it refers to a naming structure or a nominal structure
- Frequency distributions are used to measure Nominal scales
- The Mode is the statistic used for Nominal scales

Examples pertaining to the Nominal Scale:

- Sex: male or female
- Local council name: Attard, Valletta, Zurrieq
- Nationality: Maltese, Italian, Greek
- Citizenship: Maltese, Non-Maltese, Foreign
- Country: Malta, China, Antartica

[2] A very good reference on these terms can be found under: http://davidmlane.com/hyperstat/A30028.html)

The Bright Spark Memstore for Nominal:

IMAGINE NOMINAL – IMAGINE NAME

Ordinal scale

The ordinal scale serves the purpose of building up a series of numbers that are ordered in scope, however the number that is larger than another number may not necessarily be followed by another number with the same difference between the first two numbers. Thus an ordinal number that has a designation of 4 may not be bigger than 3 and 2 could be bigger than 1.

There are various characteristics that describe this scale.

- Quantitative information can be employed within this scale
- The scale refers to amounts (numbers) but not as an actual amount but as a representation of an amount
- One item in the variable does not necessarily signify the same difference between a sequence of numbers
    - Aims to discern who is 1st, 2nd, 3rd in recidivist cases
- Indicates Rank Order
    - Score of 1 means the most or the least
    - Score of 2 means the 2nd most or 2nd least
    - Example: first timer, recidivist, twice recidivist...
- The reference to order is paramount, thus one figure must still follow the next as they may signify severity of an issue such as: (1 – not important at all, 2 - low importance, 3 - important, 4 - highly important, 5 – extremely important)
- There is no true 0 (as 0 does not represent a number but a state) and the zero is designated arbitrarily
- The Median is the statistic used for Ordinal scales

Examples pertaining to the Ordinal Scale:

- Affirmation: Yes, No
- Age Cohorts (years): 0-4, 5-9, 10-14
- Density (population per square kilometre): (0, 1-100, 101-1,000, 1,001 – 2,000)
- Mohs Hardness Scale (1, 3, 9, 21, 48. 72, 100, 200, 400, 1,600

As a good indicator, one can use the Mohs Scale for Hardness. The actual scale refers to the differences between the numbers representing the different levels within the scale while each represents the relative actual hardness.

For example, the difference in hardness between scales 8 and 9 and that of 9 and 10 reflects an actual hardness difference of 200x and 400x respectively (compared to the previous number). Thus, though the intervals are the same (8-9 and 9-10 = interval = 1), in reality the hardness difference is major (200-400 and 400-1,600).

(http://www.galleries.com/minerals/hardness.htm)

Mohs Hardness Scale	**Actual Relative Hardness**
1. Talc	**1** Talc
2. Gypsum	**3** Gypsum
3. Calcite	**9** Calcite
4. Fluorite	**21** Fluorite
5. Apatite	**48** Apatite
6. Orthoclase	**72** Orthoclase
7. Quartz	**100** Quartz
8. Topaz	**200** Topaz
9. Corundum (ruby and sapphire)	**400** Corundum
10. Diamond	**1,600** Diamond

The Bright Spark Memstore for Ordinal:

IMAGINE ORDINAL – IMAGINE ORDERED

Interval scale

The Interval scale serves the purpose of building up over the ordinal in more detailed precision levels. A series of numbers is ordered but a number that is larger than another number is followed by another number with the same difference between the previous and following numbers. Thus an Interval number that has a designation of 4 has the same interval difference from 3 as 2 has from 1. However one must keep in mind that 50 is not necessarily twice 25, such as that designated in temperature measurements.

There are various characteristics that describe this scale.

- Quantitative information can be employed within this scale
- The scale refers to amounts (numbers) but not as an actual amount but as a representation of an amount
- Rarely used in social sciences
- Numbers are separated by the same gap or interval
  - Difference between 3 and 4 is the same as between 1 and 2
- There is no true 0 but 0 here represents itself as simply another number
- There can be negative numbers included such as (Negative bank accounts = EUR-1000)
- The Mean is the statistic used for Interval scales

Examples pertaining to the Interval Scale:

- Income: -EUR1,000, -EUR500, -EUR0, EUR500, EUR1,000
- Temperature in Degrees Celsius: -100^{o}C, -50^{o}C, 0^{o}C, 50^{o}C, 100^{o}C

The Bright Spark Memstore for Interval:

IMAGINE INTERVAL – IMAGINE EQUAL INTERVALS – IMAGINE CINEMA INTERVALS

Ratio scale

The Ratio scale has the same structure as the Interval Scale but zero means literally **zero**. The difference between the numbers is the same as in the Interval Scale but there cannot be any negative numbers.

There are various characteristics that describe this scale.

- Quantitative information can be employed within this scale at the highest level of precision
- The scale refer to actual and true amounts (numbers)
- Numbers are separated by the same gap or Ratio
  - Difference between 3 and 4 is the same as between 1 and 2
- There is a true 0 as it represents the base number
- 10 is exactly twice 5, 2 is twice 1
- There cannot be negative numbers (example, one can weigh **30 kilos not -3 kilos)**
- The Mean is the statistic used for Ratio scales

Examples pertaining to the Ratio Scale:

- Temperature in Kelvin: Absolute 0 (0K, 100K, 273K (relative to 0^{o}C), 1,000K
- Age in Years: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

– Richter Scale (Earthquakes Scale where each number signifies a severity that is 10 time as sever than the previous number): 0 (no earth movement), 1, 2, 3, 4, 5, 6, 7, 8, 9,

The Bright Spark Memstore for Ratio:

IMAGINE RATIO – IMAGINE RATIOS

The measurement scales need to be employed within the Matrix as described in Chapter 4. Based on the relative scale, each variable can be given a structure within which to operate. Further description of how to carry out this process will be tackled in the following chapters.

In conclusion, this chapter has enabled the reader to progress through the processes required to make the jump from an idea onto a tangible structure known as a variable. It has also helped the reader to understand how these variables can fit within the various research processes described in previous chapters.

The concept now has something which can be measured. It also has a foundation for measuring that variable through an understanding of the scales that one can use to carry out the analysis. The Matrix can now be completed and the research can be set in motion.

Prior to this launch, however, one must understand issues pertaining to data acquisition and the data quality, to ensure that the type of data is actually 'clean' enough for statistical purposes.

Questions (refer to Appendix for the answers)

1. Very briefly mention the steps between the conception of a research idea and the formulation of an indicators' list.
2. What do you understand by the research rule of Pierre Gallois?
3. What is "conceptualisation" in research?
4. What is a research "entity" and what does "entitation" in research describe?
5. What do you understand by "quantification" in relation to research?
6. After coming up with a research idea/concept, transforming it into an entity and giving it a measurement structure, the researcher still has to do something else. What is it?
7. Why does the researcher need to set up a list of indicators and a lineage process?
8. Why are indicator lists important?
9. Define a surrogate variable.
10. Why is the process of "composition" (in research) important?
11. List the four main types of measurement scales.