

# Covid-19 and Data Quality

War has always been related to life or death, and today it is clear that the combat towards Covid-19 is a global war. War and crises always call for innovation on all possible fronts, and it is a proven fact that he who has the best technology, usually win the war. The digital age led to cyber warfare, and with the latest technological advances, ML and AI plays a vital role. On 3 April IOL Tech News quoted Prof Louis Fourie: "Since AI needs large amounts of data, the challenge with Covid-19 currently is the availability of quality reliable data ".

Matter of the fact is that quality data is essential for all strategies and decisions concerning the combat of Covid-19 such as:

- Lock down at what stage or not.
- Success of preventative measures.
- Effectiveness of pre-cautions, natural remedies and assistance to the body immune system.
- Prediction on spread.
- Preparation for hospitalisation.
- Requirement for vaccine or not.
- Economic impact
- Balance between saving covid-19 lives and starvation
- Etc

The question is whether vital decisions are backed by reliable and quality data? Are the latest advances in data science and quality data life cycle being applied? Crises like the current covid-19 will not disappear and will continuously mutate to next dimensions. Whether man made or not. Hence the requirement for application of the latest technology and standards supporting digitalization and data life cycle management is essential.

## Standards

Key to global and international co-operation is standards and standardisation. World war2, initiated the NATO Codification system in the mid 1950's for data standards in a common stock identification system throughout the NATO alliance. It includes medical equipment, pharmaceuticals and drugs.

Based on similar principles a large number of codification schemes and authorities was established in various countries over the past years and is extended for standardization and control in the medical industry today.

A good example is the Health Level Seven International (HL7) organization, founded in 1987 and dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services.

Typical codes and standards developed by various health organisations are for example:

- The World Health Organization (WHO), the body responsible for publishing the International Classification of Diseases.
- The National Centre for Health Statistics (NCHS) received permission from the World Health Organization (WHO) to create the International Coding System (ICD-10-PCS) to classify

diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances and external causes of injury.

- Diagnosis Code Sets (Dx)
  - Diagnosis Related Groups (DRG)
  - Procedure Codes Sets (Tx)
  - Etc
- Current Procedural Terminology (CPT) code set is maintained by the American Medical Association. The CPT code set accurately describes medical, surgical, and diagnostic services and is designed to communicate uniform information about medical services and procedures among physicians, coders, patients, accreditation organizations, and payers for administrative, financial, and analytical purposes.
- The National Drug Code (NDC), which serves as a universal product identifier for human drugs.
- Hierarchical Condition Categories (HCCs) are used to capture medical status and history in many risk models.
- Diagnosis-related group (DRG) is a system to classify hospital cases into one of approximately 500 groups, also referred to as DRGs, expected to have similar hospital resource use.
- There is more than one DRG system being used in the United States only, not to speak of other similar systems used in other countries.
- In South Africa the National Pharmaceutical Product Index (NAPPI) code is a unique coding system for medicines, surgical or consumable products and medical procedures, which allows for a customer to claim a refund from their medical aid.

These codes and standards drive the systems which are used to manage and monitor Covid-19 Today. For example:

- Diagnosis Related Groups (DRGs) are assigned by a "grouper" program which gathers claim information based on ICD diagnoses, procedures, age, sex, discharge status and the presence of complications or comorbidities. All these factors are used to determine the appropriate DRG on a case by case basis.
- The Inpatient Prospective Payment System (IPPS) is a complex calculation which begins with each case being categorized into a diagnosis-related group (DRG). Each DRG has a payment weight assigned to it.
- Similar but different systems of categorisation and payment exists in various countries.
- In South Africa the system is driven by the National Pharmaceutical Product Index (NAPPI) codes.

The initial design of these standards and systems are noble within their silo intent. However, even within these silo data capturing systems, implementation practise yield many data quality problems such as:

- Age and Diagnosis (incompatibility)
- Age and Procedure (incompatibility)
- Gender and Diagnosis (incompatibility)
- Gender and Procedure (incompatibility)
- Gender and clinical specialty (incompatibility)
- Drug and Diagnosis (incompatibility)

- Drug and Procedure (incompatibility)
- Drug and Interaction Drug (incompatibility)
- Valid Lab Value (value out of range)
- Delta Lab Value (value out of range)
- Observation data elements (value out of range)
- Demographics data elements (value out of range)
- Time Sequence (Date and time error)
- Drug monitoring (incompleteness)
- Drug and Lab (incompleteness)
- Diagnosis and Lab (incompleteness)
- Drug in same class at same time (duplication)
- Age and Drug (incompatibility)
- Gender and Drug (incompatibility)
- In Patient Only (IPO) Procedure (incompatibility)
- Drug and Allergy to Drug (incompatibility)
- Death Date/Indicator (incompatibility)
- Tobacco user/Smokeless indicators (incompatibility)
- Alcohol user/usage indicators (incompatibility)
- Drug Dose (value out of range)

The typical data quality examples above illustrates the data quality challenges in today's isolated digital health systems. Not even to mention the global explosion of Covid-19 and the emphasis it is putting on correct diagnosis, testing, interpretation of results, correct procedure followed, treatment and dismissal on an international and global basis.

### **Data Quality Thinking**

Quality problems were experienced in the manufacturing industry after World War II. It was quickly realized that quality cannot be inspected into products and services after manufacturing, but need to be done real time within the manufacturing process. The concept of quality circles as originally described by W. Edwards Deming in the 1950s, was adopted by Toyota and the success being extended to other industries. This led the way for the development of process quality standards, such as ISO 9000. Quality needs to be part of the process and cannot be inspected into the product afterwards.

Since the rapid adoption of digital management systems in all industries, including health care from the mid 1980's, it was realized that data was key to management information. It triggered research initiatives at MIT whereby data was treated as an asset with a life cycle. The dot-com era of the late 1990's and adoption of e-Commerce since 2000, urgently required digital catalogues describing products and services in all industry sectors including healthcare.

In 2005 an initiative was launched by US Defence Logistic Information Services (DLIS) for the modernization of the NATO codification system towards an e-Commerce digital system for characteristic data exchange. This task was taken up by an ISO workgroup, ISO TC 184/SC4 which is responsible for Industrial Data. The ISO 8000 standard for data quality was initiated to ensure data quality in inter alia ISO 9000 industrial processes. The main aim is the creation of data quality standards to ensure digital data portability between various stakeholders in a specific value chain, whether between the departmental data silos in organizations, between organizations or countries.

South Africa is one of the founder members of ISO 8000 and a mirror committee in the South African Bureau of Standards, SABS TC 184 was established in 2007.

Key to ISO 8000 is master data which need to be portable for data exchange between industries, including the Health Industry. SABS TC 184 has the membership participants to facilitate data quality for Covid-19.

### **Master Data**

Master data is a type of data that describes subjects related to the 'who,' 'what,' and 'where' in transactions, communications, and events in the combat to covid-19 for example. The 'who' could be a customer, employee or potential virus carrier; the 'what' could be the virus condition, a product or service, and the 'where' could be a residential address, store, office, or a virtual location.

Management and governance of master data related to Covid-19 is at the root of data science, trusted scientific analysis supporting informed decisions and the establishment and formulation of Covid-19 combat strategies. ISO 8000 data quality frameworks, templates and processes are described to ensure governed data management in both structured data environments (such as described in the health code management hierarchies) or in the unstructured big data environments (such as social media). The challenge currently is to apply these latest data science technologies, to the regional and global monitoring and generation of data sets, which come from dispersed locations, different sources and different jurisdictions and countries. Not even to mention the different codification systems, localised standards and languages.

These big data sets need to be scientifically gathered, analysed, harmonised, data quality checked and modelled for integrated system interaction. This requires a dynamic systems approach enabling the integration of various disciplines, services and technologies in the combat of crises such as covid-19.

### **4<sup>th</sup> Industrial Revolution Digitalization Technologies to be applied**

Maybe magic is happening behind the scenes, but in the current global Covid-19 combat, little to no evidence is illustrated for a holistic approach of data management practices, utilizing the almost daily enhancements in the digital technologies of the 4<sup>th</sup> industrial revolution. Essential data quality methods need to be integrated into the cyber and digital technologies deployed in the health industry such as:

- Big Data and Identity Resolution
- Codified and structured data integrated with social media data.
- Machine learning and artificial intelligence supported by correct and governed data sets.
- Augmented algorithms for artificial components, machine and human interaction.
- Taxonomies, dictionary and ontologies for integrated health industry.
- Master data based on Quality Identifiers and Authoritative Legal Entity Identifiers.
- Schemas and frameworks for digital data exchange.
- Frameworks to govern master data.
- Data conforming to POPI (GDPR).
- Cloud hosted and Data Quality Hub integrations.
- Etc.

## Conclusion

On 29<sup>th</sup> April 2020, after one month of quarantine, the total statistics for South Africa provinces was:

| Province      | Positive Cases | Recoveries | Deaths  | New Cases | Tested  |
|---------------|----------------|------------|---------|-----------|---------|
| Gauteng       | 1408           | 897        | 11      |           |         |
| Limpopo       | 31             | 25         | 2       |           |         |
| Mpumalanga    | 31             | 18         | 0       |           |         |
| Kwazulu-Natal | 956            | 364        | 32      |           |         |
| Free State    | 113            | 92         | 5       |           |         |
| North West    | 29             | 16         | 0       |           |         |
| Northern Cape | 17             | 10         | 0       |           |         |
| Eastern Cape  | 630            | 46         | 11      |           |         |
| Western Cape  | 2135           | 605        | 42      |           |         |
| Total         | 5350           | 2073       | 103     | 354       | 197127  |
| Population %  | 0.0089%        | 0.0035%    | 0.0002% | 0.0006%   | 0.3285% |

The one serious question to ask in view of the above statistics and context above?

Is there enough evidence and can we trust the data for the harsh decisions for lock down and continuation of lock down. Especially within the context of poverty and economy halt!

Or is there some serious global manipulation and hidden power at work? Above governments with fear manipulation.

Are we indeed blinded by manipulated and controlled media? Why are only pharmaceutical and vaccine solutions for Covid-19 propagated and no alternative break troughs such as negative ion confirmation for virus destruction. Only to mention one.

<https://youtu.be/NNxNKIbPuHM>

What append since our historical evidence of more than 2000 years ago, when incurable diseases was demonstrated to be eradicated by believe and spiritual energy. Even greater miracles and defying of death as well as a promise that human kind is capable of more if the right choices of love and not fear is applied.

Is it not time that our science, research and quality data evidence be focused on vibrational and spiritual sciences compared to selfish, ego centric, material money chasing and manipulation.

Dr Salomon de Jager Pr. Eng.

Founder and Director Pilog Academy.