

iq International Journal

VOLUME 13 | ISSUE 1 | APRIL 2017

Information Quality Journal

A Publication of
IQ International

iaidq.org

Aligning Data Quality and Information Quality using Data Standards

by
Salomon de Jager
Inca Vtipil

Introduction

As digital technology advanced during the 80's and early 90's and digital information management systems became the norm, the issue of system integration and data warehousing became a critical part of rendering business value for data systems. It quickly became clear that data warehousing was capable of contributing to Data and Information Quality problems without proper oversight and management. To preventing data quality problems and allow for the management of information as an asset with processes, Total Quality Management (TQM) (and later Total Quality Data Management (TDQM)) was propagated by MIT, as a policy to treat data and information as an Enterprise Asset. Currently, this is evolving into a policy of Data Governance under the position of Chief Data Officer (CDO) in businesses and enterprises worldwide.

The ISO 8000 Data Quality standards, initiated in 2005, focused on supporting data dimensions in the quality processes advocated in ISO 9000 and ISO 9001. The critical role of Master Data in Data Quality business processes was identified and addressed in the ISO 8000 part 100 series. Discussions and debate on the understanding of data and Data Quality vs Information Quality remains an interesting topic, especially when Data and Information Quality measurement and Return on Investment (ROI) are supported as a means of putting data quality programs into action. Processes and change management to ensure the successful implementation through the total organization, including Data and Information Quality ownership, remains a challenge.

This article put emphasis on the deeper understanding of data, information, and how standards can support management leadership to achieve Master Data Management. This paper supports matching framework to language and employee needs by leveraging ontology principles underlined by the ISO 8000 part 100 series.

Data and information management systems are an essential and integral part of the modern business landscape.

Human Needs and the Data Hierarchy

In 1943, psychologist Abraham Maslow introduced his theory of the Hierarchy of Human Needs, in which users seek to address fundamental physiological needs first, and then progress upward through higher-level needs for safety, then love, esteem, and, finally, self-actualization (1). This theory of human motivation is depicted in the shape of a pyramid with the most fundamental physiological needs at the bottom and increasingly higher-order needs built on top. At the apex is "self-actualization," a psychological state in which an individual has fulfilled each previous level of need (food/shelter/safety, love, esteem) and is able to reach his or her full potential.

Bostick in his book, "Ascending the Data Infrastructure Hierarchy", uses the triangle framework to investigate data infrastructure needs (2). His equivalent hierarchy renames the five stages as Tribal, Enforced, Standardized, Actualized and Peak performance.

In view of the rapid increase of daily data and the volumes of unstructured data foreseen in the near future, business will need new ways to manage data of all types with workflow and processes to turn raw data into quality actionable information. One can imagine a similar hierarchy of needs for Data – beginning with data creation and storage at the foundation levels, reaching to actionable information in distributed data centers ensuring business success. Taori and Caroyano suggest in “Data as a Service and the Analytics Hierarchy of Needs” the following levels in the Data as a Service (DaaS) hierarchy (3). Refer to Figure 1 below.

Quality Data interpreted pragmatically can become actionable information, which in turn can lead to Knowledge, Intelligence, and ultimately to Wisdom. Quality Data therefore can be depicted in a similar hierarchical way, where the apex of Wisdom can complement the higher levels of Human Needs such as “self-actualization”. Figure 2 depict this similarity:

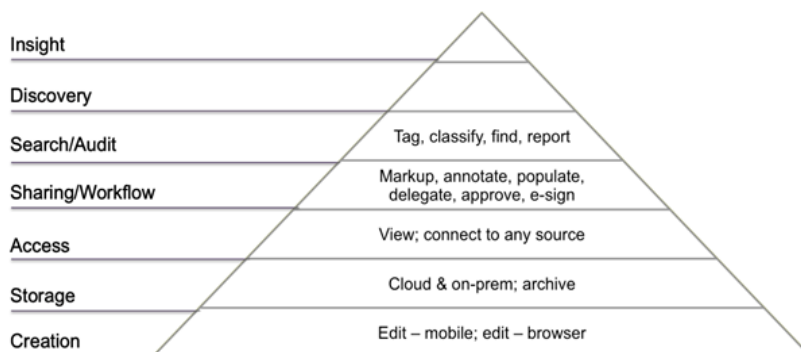


Figure 1: Data as a Service (DaaS) Hierarchy

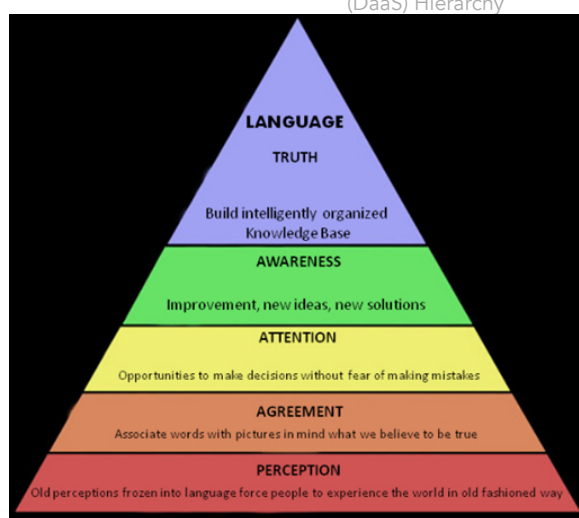
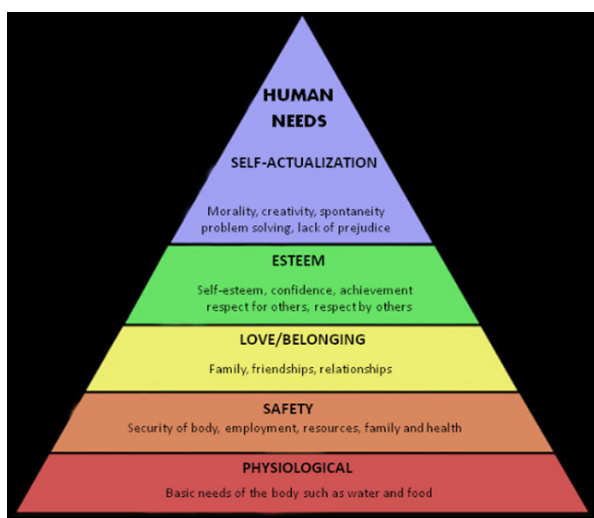


Figure 2: Similarity between Human Needs Hierarchy and Data Hierarchy

Data Hierarchy and Language Hierarchy

Human needs, which essentially manifest in the human brain and mind, is mostly transferred to the Data Hierarchy through Language. In a similar way a matching hierarchical parallel can be drawn between the Data Pyramid (Data, Information, Knowledge, Intelligence, Wisdom) and the developed Language Pyramid (Perception, Agreement, Attention, Awareness, Truth). Figure 3 depict this similarity:

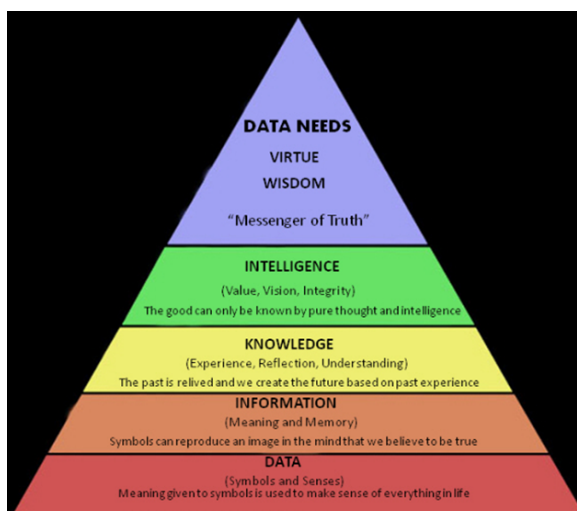


Figure 3: Similarity between Data Hierarchy and Language Hierarchy.

The relationships and similarities between the Data and Language hierarchies is defined and explained below. It is necessary to have this understanding as it enables the construction of an Ontology that will be defined and supported by our Standards.

In the new oxford dictionary, data is defined as: “facts and statistics collected together for reference and analysis”. Data is associated with syntax, otherwise it would not be machine readable. In another way Data can also be regarded as Symbols and Senses.

What we call facts can vary from person to person, based on individual perception. Perception can be defined as the way in which something is regarded, understood or interpreted. The word “Collar” might mean the collar of a shirt, a collar to lead a dog, or a mechanical collar to a bearing, for example. Data is based on the individual perception of a word. A collection of data for which there is no relation between the pieces of data is not information. Data should not be confused with Information. Information has reference to meaning and memory. Information relates to description, definition, or perspective and may originate from data.

We need to reach agreement in understanding the relationships among data. The problem is that we associate symbols, words, and pictures in the mind of what we believe is true. The solution is that we need to reach agreement on the understanding of the words in question. This requires semantics and semantic encoding.

Beliefs tell us how something may be. Knowledge tells us how something is. Learned knowledge is distilled out of beliefs. It has to do with experience, reflection and understanding. Knowledge is not about what we know, but what we don't know and need to find out or discover. Knowledge is comprised of strategy, practice, method or approach. It helps to focus attention and is about possibilities and making decisions without fear of making mistakes. We need to change the power of belief we vest in symbols.

Intelligence is a process or innate capacity to use pure knowledge in order to respond to ever-changing requirements. It is the ability to perceive information and retain it as knowledge to be applied as an adaptive behavior within the environment. In the mental capacity it involves the capability to reason, plan, think abstractly, resolve problems, comprehend complex ideas, and learn from experience, and reflects a deeper capability of comprehending our environment.

Through pure thought and intelligence, awareness is achieved. Awareness is the ability to directly know, perceive, feel, be conscience of events, objects, thoughts or emotions, all without necessarily implying understanding. Awareness leads to new improvements, new ideas, and new solutions.

By balancing the adaption of intelligence and awareness, wisdom is achieved. Wisdom determines how an organization grows and flows. wisdom is the ability to having experiences, knowledge, good judgment, awareness and understanding – i.e. the quality of being wise. Wisdom arises out of the foundational principle of intelligence and how it is understood. In short: the ability to adapt by looking for real truth, or to base decisions out of that what is true or in accordance with fact or reality.

Using Data and Language Hierarchies to Satisfy Business and Human Needs

It is well-known that Data and Information Quality do not happen merely by themselves; they are accompanied by dedicated management processes and effort. The hierarchical relationships between human requirements and data, as described above, can only be achieved via language facilitation. It entails a spiral process whereby the lower level elements need to be implemented and understood before the next level can be successfully implemented or achieved. The driving force in the progress of Data and Information Quality to next levels in the hierarchy of requirements, is achieved by leadership, facilitated by language and ontology. Refer to figure 4 below:

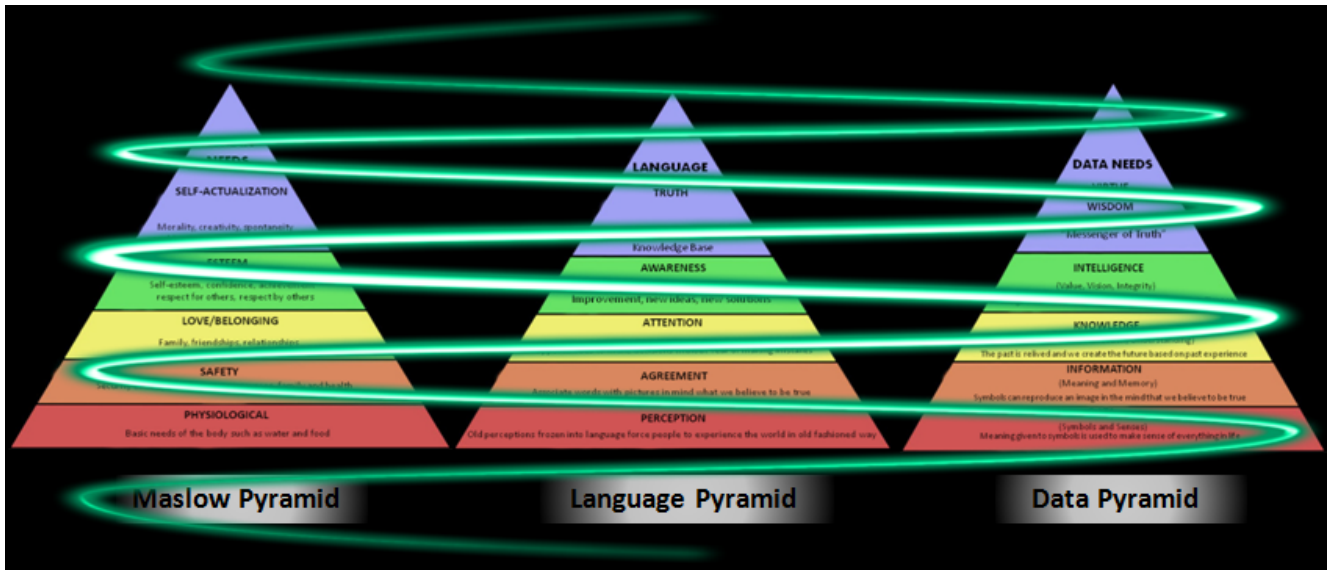


Figure 4: The spiral management process facilitated by language to ensure Data and Information Quality meet the hierarchical human needs as well as business requirements.

Leadership Hierarchy

The change management and driving force for Data and Information Quality is initiated by bold leadership. The equivalent leadership hierarchy consists of the levels: – (Manage Perception, Art of Agreement, Channel Attention, Timeless Awareness, Destined Truth). These levels are illustrated in figure 5 below.

The upward leadership spiral (Refer Fig 4) to ensure Data and Information Quality to support human and business requirements is facilitated by the language hierarchy. Language in turn, need to be complimented by an ontology based on standards. Data standards and governance rules play a vital role in the success of a Data and Information Quality programs.

Figure5: Leadership Hierarchy as spiral catalyst to achieve human and business requirements, supported by quality data and information.



Pragmatic Framework for Data and Information Quality

Data and Information Quality are measured in terms of “meeting requirements”. Data is regarded as quality data if it conforms to the required syntax and can be interpreted within the required semantics. Quality Data and Information therefore entail syntax, semantic encoding, as well as pragmatics. To measure Quality Data and Information, the right answers need to be asked and evaluated. Getting to the right questions that will lead to the higher levels of Data and Information quality as depicted in the hierarchical discussions above is a challenge. By applying the leadership hierarchy and performing the activities of Analyze, Identify, Action, Motivate, and Serve, as well as performing quality inquiries on the needs or requirements hierarchies with the support of the language hierarchy will contribute to the Data Quality hierarchy. Asking and evaluating questions according to a systematic framework and sequence (1 to 20) as suggested in Figure 6 below, add great value to the measurement of Quality Data and Information.

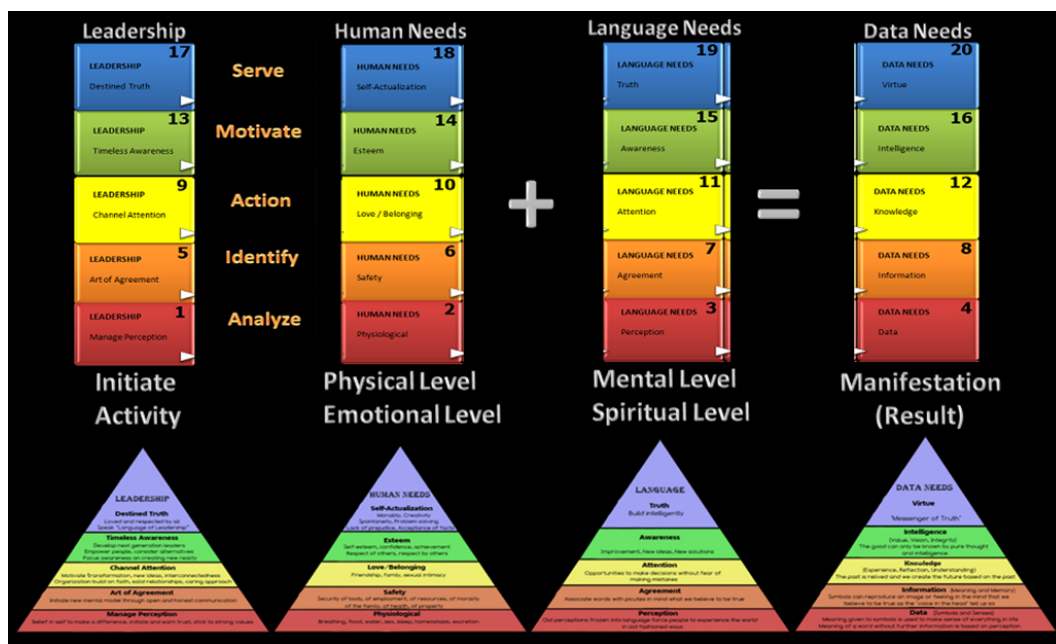


Figure 6: Framework to apply Leadership hierarchy activities to requirements (needs) and language hierarchy as support for pragmatic evaluation (questions and answers) for Data and Information Quality measurement and evaluation.

ISO 8000 Data and Information Quality

ISO 8000 is a Data Quality standard developed to support the management of data as an asset. ISO 8000 supports the quality processes as defined in ISO 9000 and ISO 9001, and was developed to support the Data and Information Quality needed in technical and business processes.

ISO 8000 fully supports the leadership spiral approach as described above, thus ensuring business requirements are met and supported by quality data and information. It is facilitated by language structured in an ISO 8000 specified ontology ensuring the fundamentals of the applicable governance rules. Refer to Figure 7 below:

ISO 8000 Master Data

Master Data Management (MDM) is as one of the key elements in ISO 8000.

The relation of Master Data to other data is illustrated below:

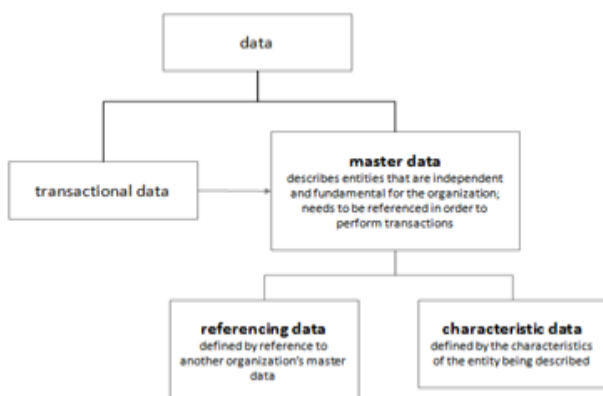


Figure 8: The relationship of Master Data to other Data

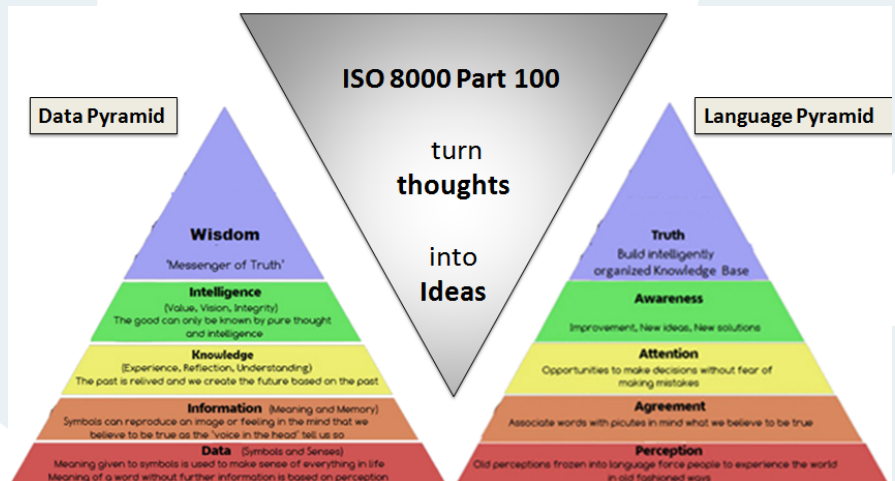


Figure 7: The spiral role of leadership and standards to progress data to wisdom with the aid of language.

ISO 8000 part 100 series is dedicated to ensuring maintainable and quality Master Data that can be exchanged independently from the software application. Part 110 deals with data syntax, semantic encoding, and Data Requirement specifications. Part 120 deals with provenance, part 130 with accuracy, and part 140 with completeness. The scope of ISO 8000 part 100 series is illustrated below:

Prerequisite for ISO 8000 part 100 (Master Data) Compliance

The main objective of ISO 8000 part 110 is to structure master data unambiguously for data portability between systems. In practical terms it means that the data must conform to a defined syntax and be semantically encoded, conforming to international recognized identifiers with entries into an open technical dictionary (OTD) for related parties. The data dictionary entries must have an identifier, term, and clear definition. A typical structure of an OTD is illustrated below:

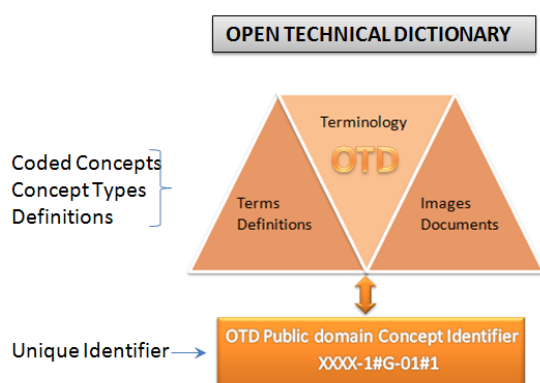


Figure 10: Example structure of Open Technical Dictionary

Each part being a standard in its own right....

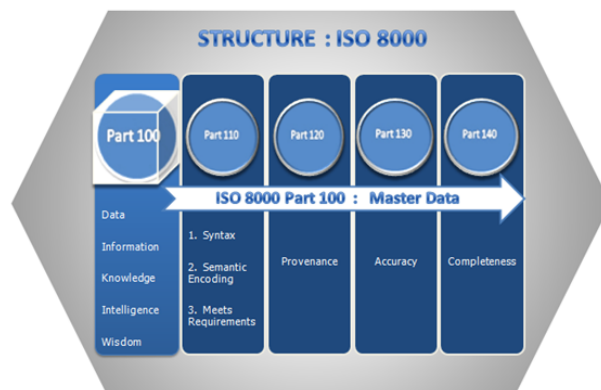


Figure 9: The structure of ISO 8000 part 100 being Master Data.

The full power of ISO 8000 part 110 is realized with the creation of coded value pairs structured into a data requirement (template) for classes of concepts, reflecting the related attributes, values, and unit of measures. By coding related metadata to the OTD and allocating classification structures, the dictionary can maintain a full taxonomy as well as ontology.

Ontology

An ontology is an explicit formal specification of how to represent the objects, concepts and other entities consistent with the definitions recorded. In information technology, an ontology is a set of concepts - such as things, events, and relations - that are specified in some way (such as specific natural language or coded as described in ISO 8000 part 110) in order to create an agreed-upon vocabulary or schema for exchanging information. The backbone of an ontology consists of a generalization or specialization hierarchy of concepts, i.e. a taxonomy.

Value of Ontology in Data Quality

Once the philosophy of data and information as an asset is adopted, as well the language and the leadership pyramids spiral, quality controls can be built into data aggregation (whether by manual entry, bulk upload or big data flows). This should be designed into the data sourcing and acquisition strategy, supported by the appropriate designed IT architecture. Mapping the aggregated data from a staging area against a well-defined Master Data Governance and management system, can only be achieved once a common taxonomy and well-managed metadata is integrated into the ontology. This also serves as a basis for cleaning and structuring data after extract, transfer, and load (ETL) processes are performed and data is transferred from one system to another. Data Governance is only possible with the existence of an enterprise dictionary for regulating and gate keeping new item entries.

The current data explosion results in a struggle for actionable information based on quality data. Quality data is very much internal focused and quality information external. Human beings are the target audience of quality information in the end, and the challenge is to bridge human thought and ideas to ensure that the correct data and actionable information is provided in support of the required hierarchical needs. An ontology backed by standards such as ISO 8000 is a necessity, and ensure human thinking is represented in a digital environment. Understanding of business requirements, facilitated by language capability, supported by an ontology in an open technical dictionary, based on standards such as ISO 8000, will ensure data and information quality within a governed IT environment.

LANGUAGE
TRUTH

Build intelligently organized Knowledge Base

AWARENESS
Improvement, new ideas, new solutions

ATTENTION
Opportunities to make decisions without fear of making mistakes

AGREEMENT
Associate words with pictures in mind what we believe to be true

PERCEPTION
Old perceptions frozen into language force people to experience the world in old fashioned ways

LEADERSHIP
HUMAN NEEDS

Build intelligently organized Knowledge Base

DATA NEEDS
VIRTUE WISDOM

"Messenger of Truth"

INTELLIGENCE
(Vision, Vision, Integrity)
The good can only be known by pure thought and intelligence

KNOWLEDGE
(Experience, Reflection, Understanding)
The part is refined and we create the future based on part experience

INFORMATION
(Measuring and Sensing)
Symbols create perception an image in the mind that we believe to be true

DATA
(Symbols and Senses)
Meaning given to symbols is created to make sense of everything in life

COLLAR, BEARING

A retaining ring designed to be mounted on a shaft in various ways and used to locate and retain a bearing.

0161-1#TM-1099479#1

References

- 

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license. To view a copy of this license visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

About the Authors



Dr. Salomon de Jager completed his Hons. B-Eng. degree (Mech. Ind.) in 1978 at the University of Stellenbosch South Africa. During his 2 year National Service in the South African Army he became an officer, contributed to Weapon Systems Development, Logistic Support System Design and obtained his M.Eng. (Ind.) degree in 1981 at the University of Pretoria South Africa. Dr. de Jager completed his PhD in System Engineering and Logistic design in 1993 at Witwatersrand University Johannesburg. Since 1985 Salomon was the head of the functional design team of Paradigm Systems. In 2014 Dr Salomon was appointed as an Affiliate Faculty member of University Arkansas Little Rock (UALR) to assist the UALR IQ program through expert advice, teaching and student supervision. Today Dr Salomon is the Chief Executive Director (President) of the Pilog Group. The Pilog Group is established from Pilog International (BVI Company) with companies in South Africa, UAE, Saudi Arabia, India, Russia and USA.



Inca Vtipil's journey with data started in the mid 80's while working for Paradigm Systems Technology as a software analyzer and tester, training officer and technical writer.

"I understand that the more data the world collect, the bigger problem is manifested due to a lack of human awareness, which is not a fuzzy philosophical flourish, but a key part of the brain's machinery for processing data.

To support an emerging shift of consciousness in the world, people are in need of real world tools and inspiration to cause positive change and support a higher vibration."