



DATAKYRA

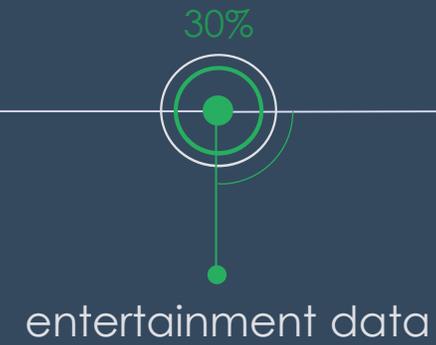
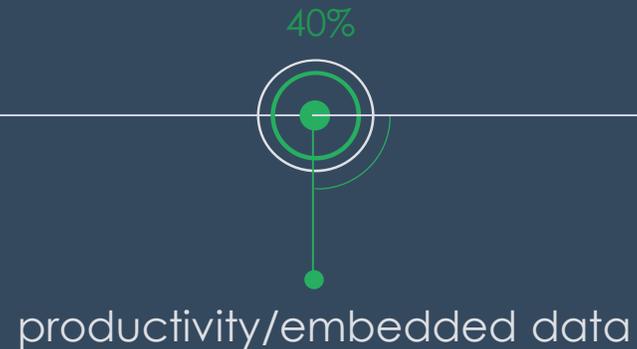
Quality of Data

and it's relevance for commercialising data

Data of next 3 years

>

Data of last 30 years



”

50% is Dark Data

66% of manager stated in study

1% of data is analyzed

IBM

“



Structured Data



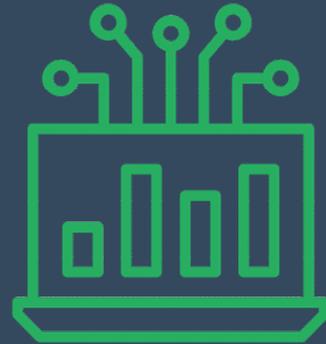
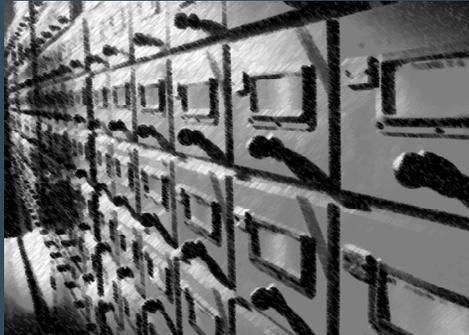
Semi-structured Data



Unstructured Data



Data Warehouse



Data Lake



Why Quality

PRECISE
NUMBER x GARBAGE = GARBAGE

$\sqrt{\text{GARBAGE}}$ = LESS BAD
GARBAGE

$(\text{GARBAGE})^2$ = WORSE
GARBAGE

$(\text{PRECISE})^{\text{NUMBER}}$ GARBAGE = MUCH WORSE
GARBAGE

xkcd.com



Quality Dimensions

Completeness

- Proportion of stored data against the potential of 100%



Uniqueness

- Record only once in regard to how it is identified (No duplication)



Timeliness

- How fast is stored information available / how old is the information



Validity

- Data matches syntax and definition



Accuracy

- Does the data describe reality



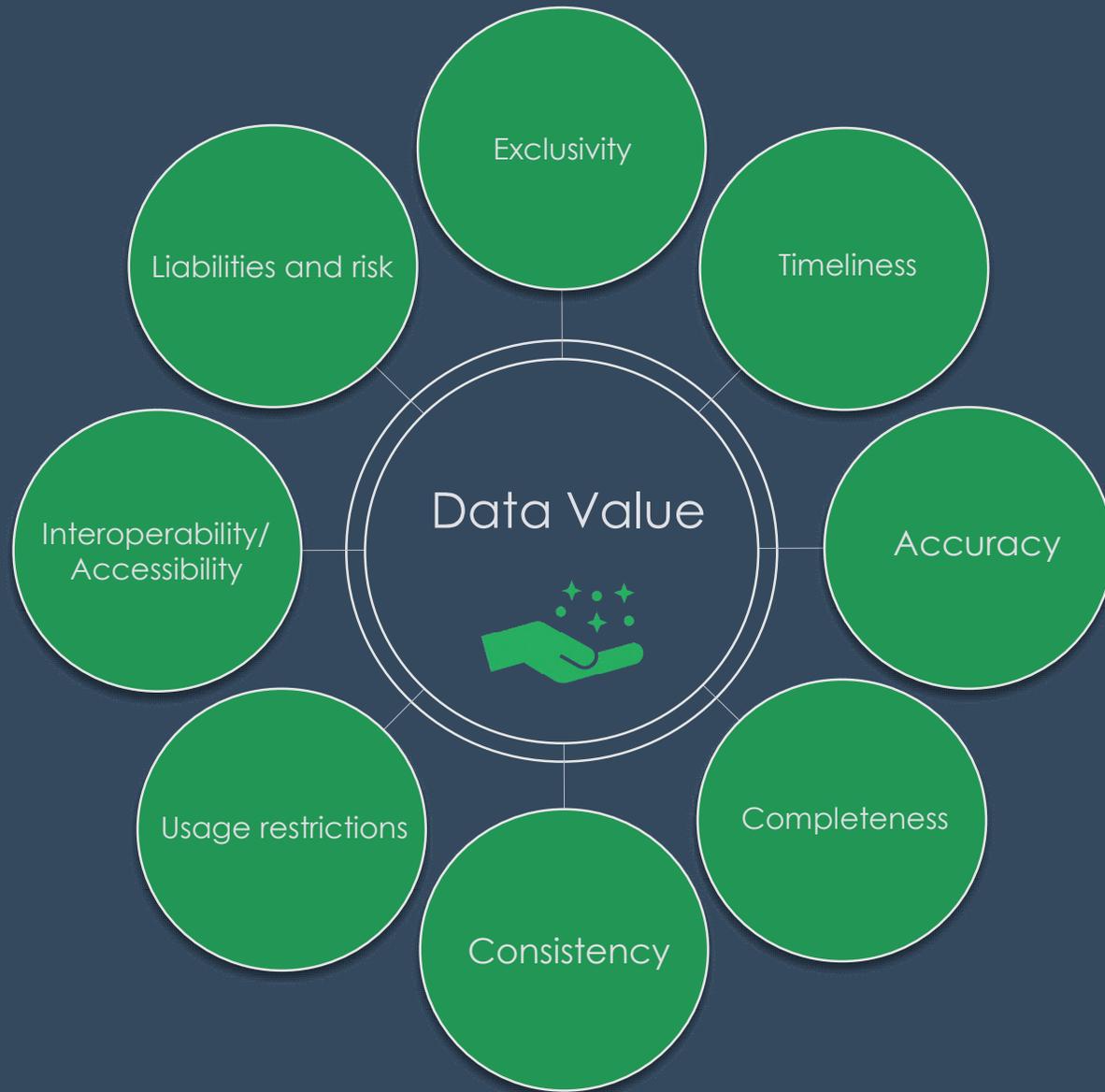
Consistency

- Does information stored in one place match relevant data stored elsewhere?



Data is Value





How to get value into data

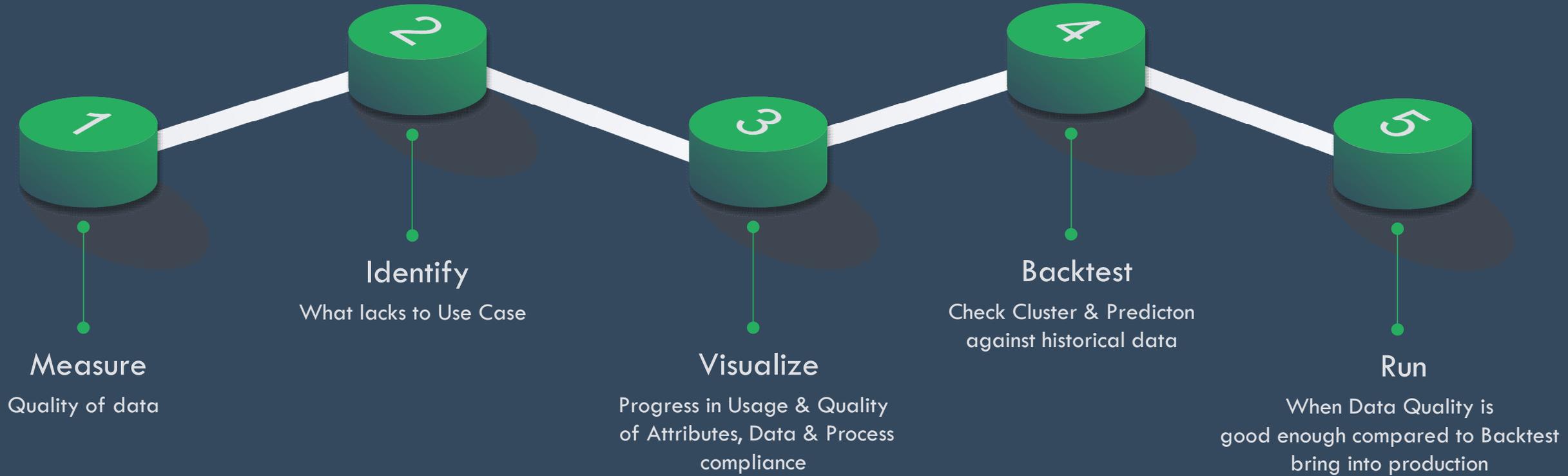
01 Quality: Fitness for use

Data Quality is a journey 02

03 Journey has to follow a precise path



Ongoing Data Quality Increase



”

FIND

what you haven't been searching for

“

DATAKYRA.com

