ADABI

UNIVERSITÀ DEGLI STUDI DI TORINO

Data Quality nell'era dei Big Data

roberto leombruni
Università di Torino e Laboratorio Revelli

ARCHIVIAZIONE, DISSEMINAZIONE E RIUSO DEI DATI: A CHE PUNTO SIAMO?

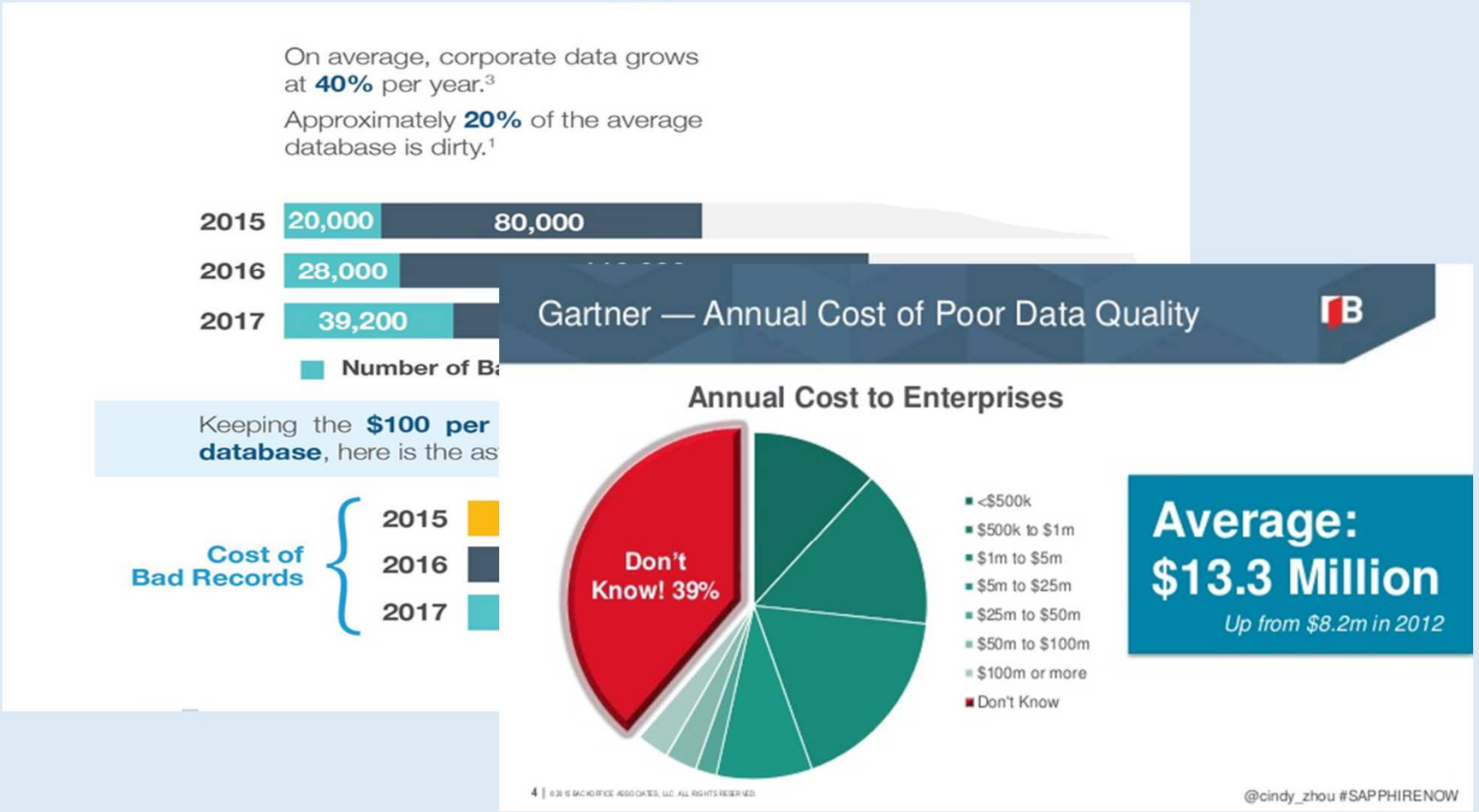
Torino
Campus Luigi Einaudi
23 novembre 2017

l'importanza della qualità

La NASA lanciò lo space shuttle *Challenger* il 28 gennaio 1986. Pochi momenti dopo il decollo, uno dei razzi propulsori esplose, portando alla distruzione dello shuttle e alla morte dei sette membri dell'equipaggio.

L'esplosione dello space shuttle Challenger fu attribuita a un problema di qualità dei dati con i quali si era gestito il decollo (Fisher & Kingma, 2001).

l'importanza della qualità



cosa intendiamo per *data quality*

«The fitness for use of information»

Martin Eppler

«The state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use»

Government of British Columbia

…

cosa intendiamo per *data quality*

rilevanza

accuratezza

completezza

consistenza

timeliness

accessibilità

comparabilità

costi

...

cosa intendiamo per *data quality*

rilevanza

accuratezza

completezza

consistenza

timeliness

accessibilità

comparabilità

costi

...

cosa intendiamo per *data quality*

rilevanza

accuratezza

completezza

consistenza

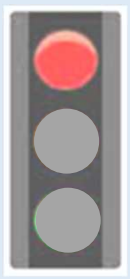
timeliness

accessibilità

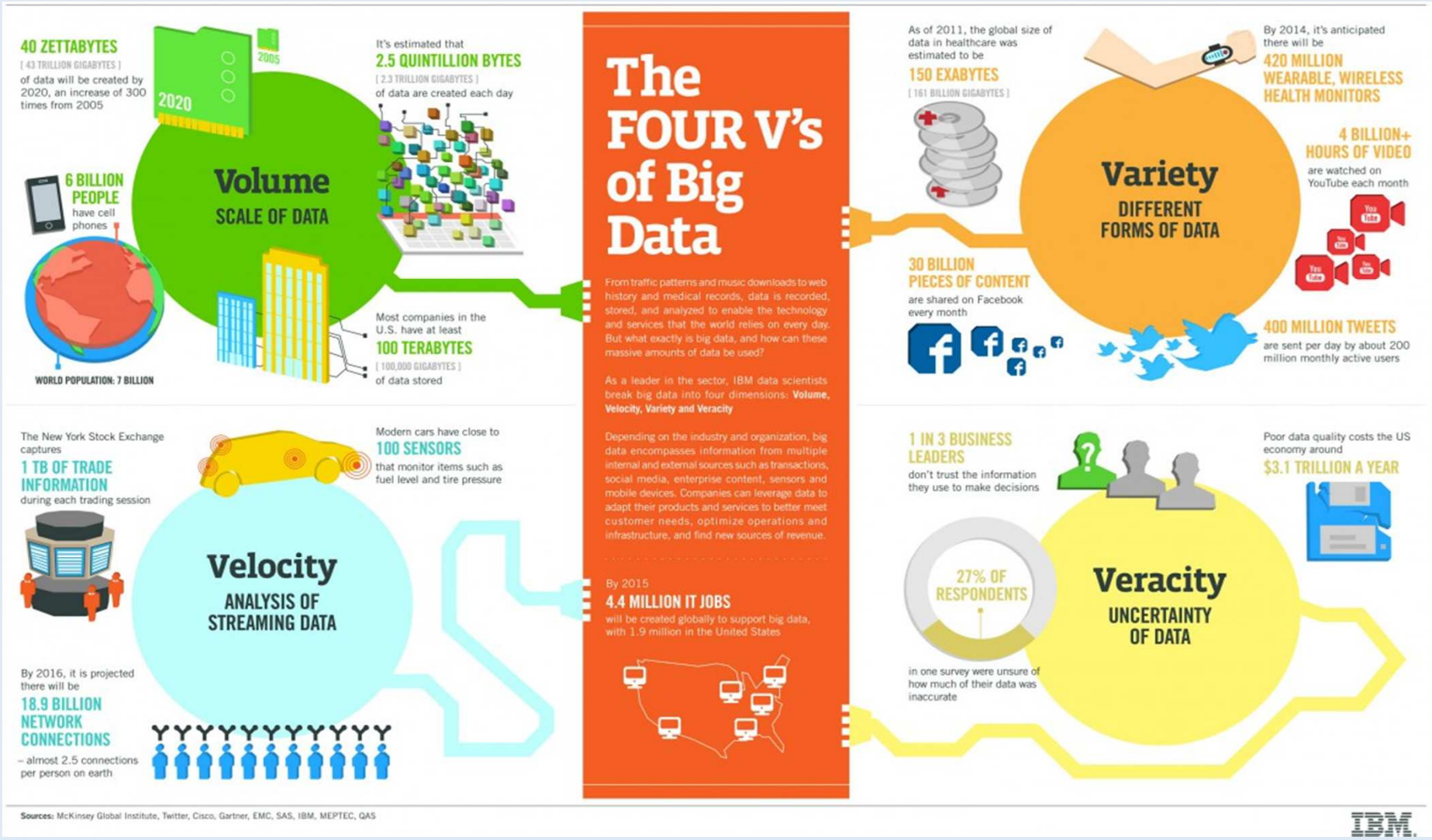
comparabilità

costi

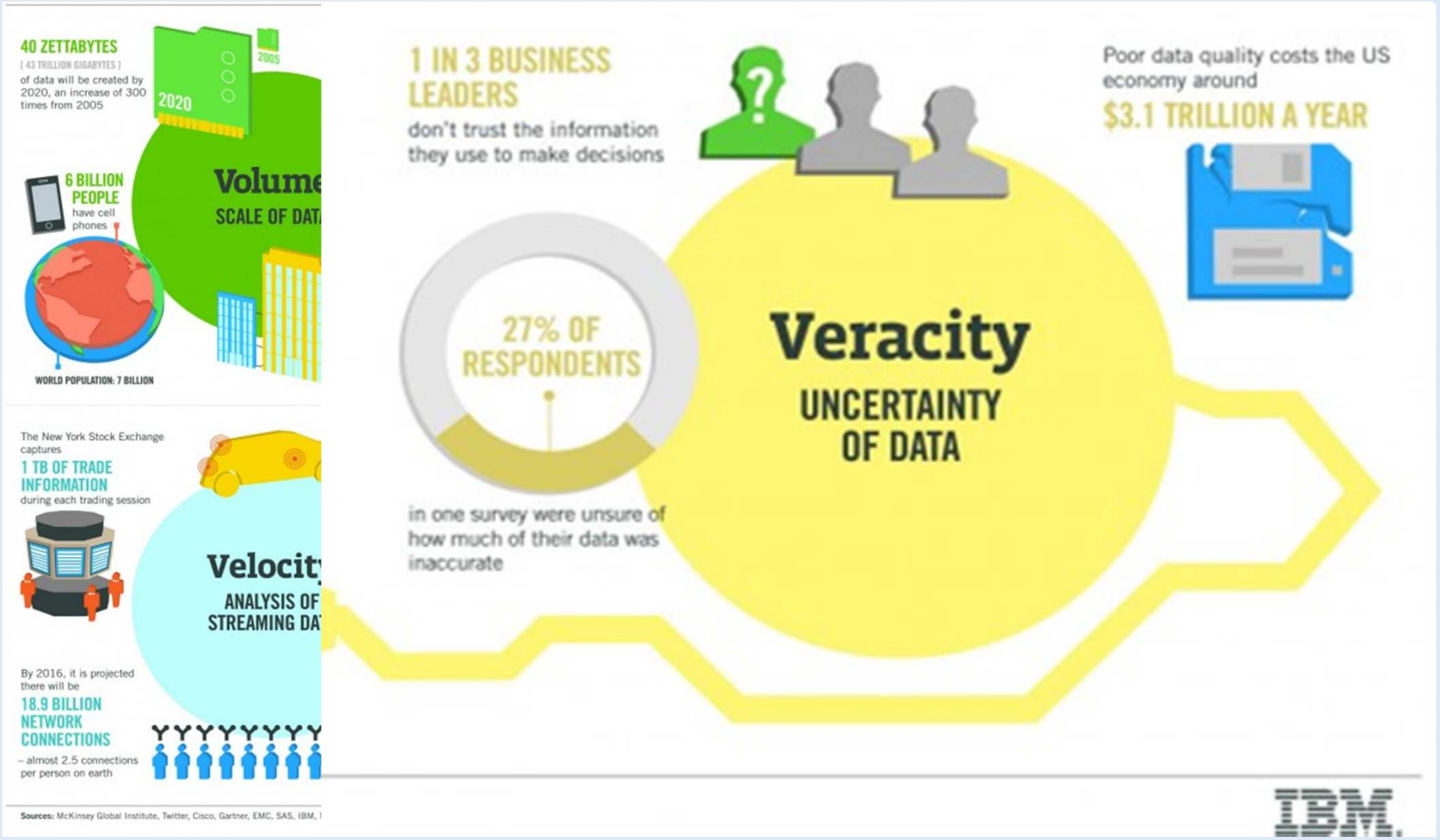
...



i costi della poca accuratezza



ancora sui costi della cattiva qualità



e gli altri costi?

Journal of Industrial Engineering and Management

doi:10.3926/jiem.2011.v4n2.p168-193

JIEM, 2011 – 4(2): 168-193 – Online ISSN: 2013-0953

Print ISSN: 2013-8423

The costs of poor data quality

Anders Haug, Frederik Zachariassen, Dennis van Liempd

University of Southern Denmark (DENMARK)

adp@sam.sdu.dk; frz@sam.sdu.dk; dvl@sam.sdu.dk

e gli altri costi?

no «fitness for use» → bad decisions

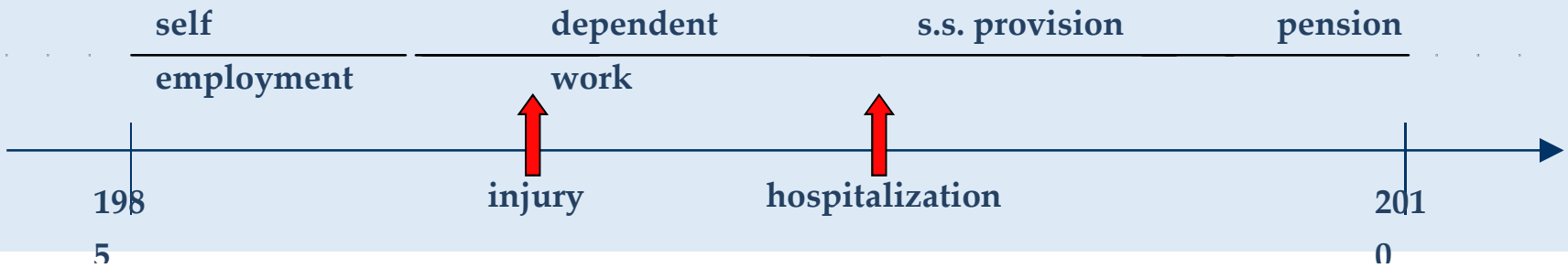
(trash in - trash out principle)

un esempio dalla ricerca sociale: WHIP-Salute

WHIP stands for *Work Histories Italian Panel*

the database is able to track the main events of individuals' working careers

It is based on **administrative data** collected by the Italian National Institute for Social Security (INPS), National Institute for Work Injuries Insurance (INAIL), Ministry of Welfare, National Institute of Statistics (ISTAT).



un esempio dalla ricerca sociale: Whip-salute