



Smart Open Data

Cerved Story

Stefano Gatti
Torino, 9 Ottobre 2014



Summary

- “ Something about me
- “ Cerved figures and market
- “ Cerved data innovation
- “ Cerved proprietary data
- “ Open Data: Cerved vision
- “ Open Data: Cerved use cases
- “ Data Quality: a strategic step in datascience
- “ Some (not definitive) thoughts about datascience
- “ Q&A

Something about me

- “ Data lover
- “ Agile organization & mindset supporter
- “ Innovation & Data Sources Manager in Cerved
- “ A runner or better an endurance sportman
- “ A knowledge sharing and open-culture passionate
- “ A nerd father of two nerd children

More about me ...

- “ Twitter: @micio1970
- “ Mail : stefano.gatti@cervedgroup.com or st.gatti@gmail.com
- “ My website: <http://www.stefanogatti.info/>
- “ My blog: <http://www.stefanogatti.info/nuvolediconoscenza/>



“Costruiamo INFORMAZIONI sulle aziende per supportare DECISIONI partendo da DATI ufficiali e ufficiosi attraverso processi tecnologici cercando di elevarLI a CONOSCENZA anche attraverso risorse umane in apprendimento continuo”

LINCE
Cerved Group

Extra Now

JULIGHT
SOCIETÀ A RESPONSABILITÀ LIMITATA

VIA CUZIO 42
27100 PAVIA (PV)

Evato il 14/05/2012
Codice Cerved Group 13095817
Servizio Richiesto: Extra Now
Situazione impresa: Attiva
Presente sul mercato dal: 21/12/2011
Richieste informazioni ultimi 6 mesi: 1
REG.CODICE 13152584

Sintesi analista

Richiedi il servizio Top Business per avere una valutazione approfondita da parte di un analista finanziario... [clicca qui](#)

Focus

| | |
|-------------------------------|------------------------------|
| Segnalazioni Rilevanti | |
| Operazioni Straordinarie | SI |
| Eventi Negativi | Nessuno |
| Ritagli Stampa Negativi | Nessuno |
| Risultato analisi | |
| Appartenenza ad un gruppo | No |
| Andamento settore | Rischio inferiore alla Media |

Cerved Business Areas



**CREDIT
INFORMATION**



**MARKETING
SOLUTIONS**



**CREDIT
MANAGEMENT**

1000 report/min

✓ Document and data search

2 million

✓ Credit scoring reports

450,000

✓ Private credit ratings

51 million

✓ Payment transactions recorded

160,000

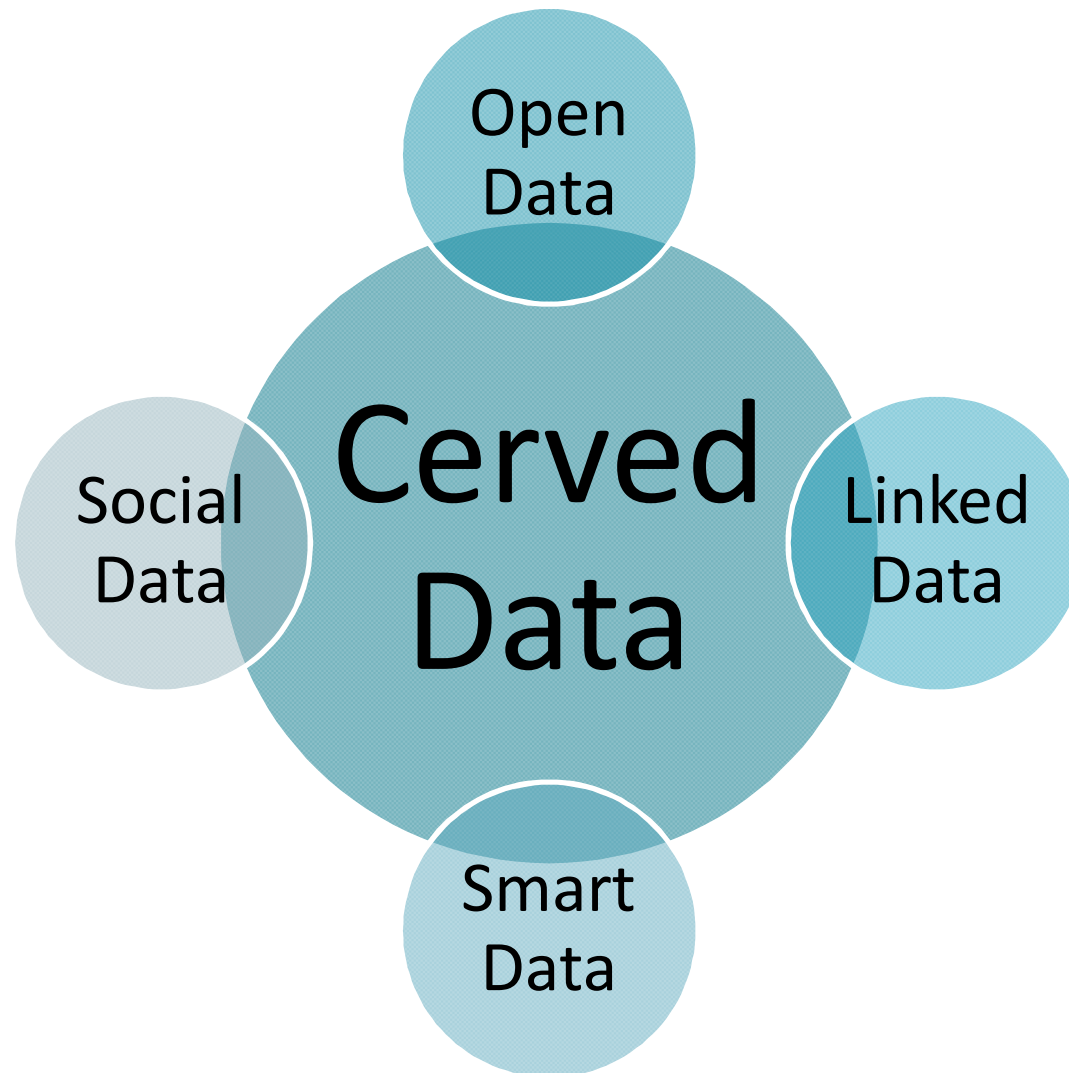
✓ Italian group analysed

313 million Euro (2013)

✓ Revenue

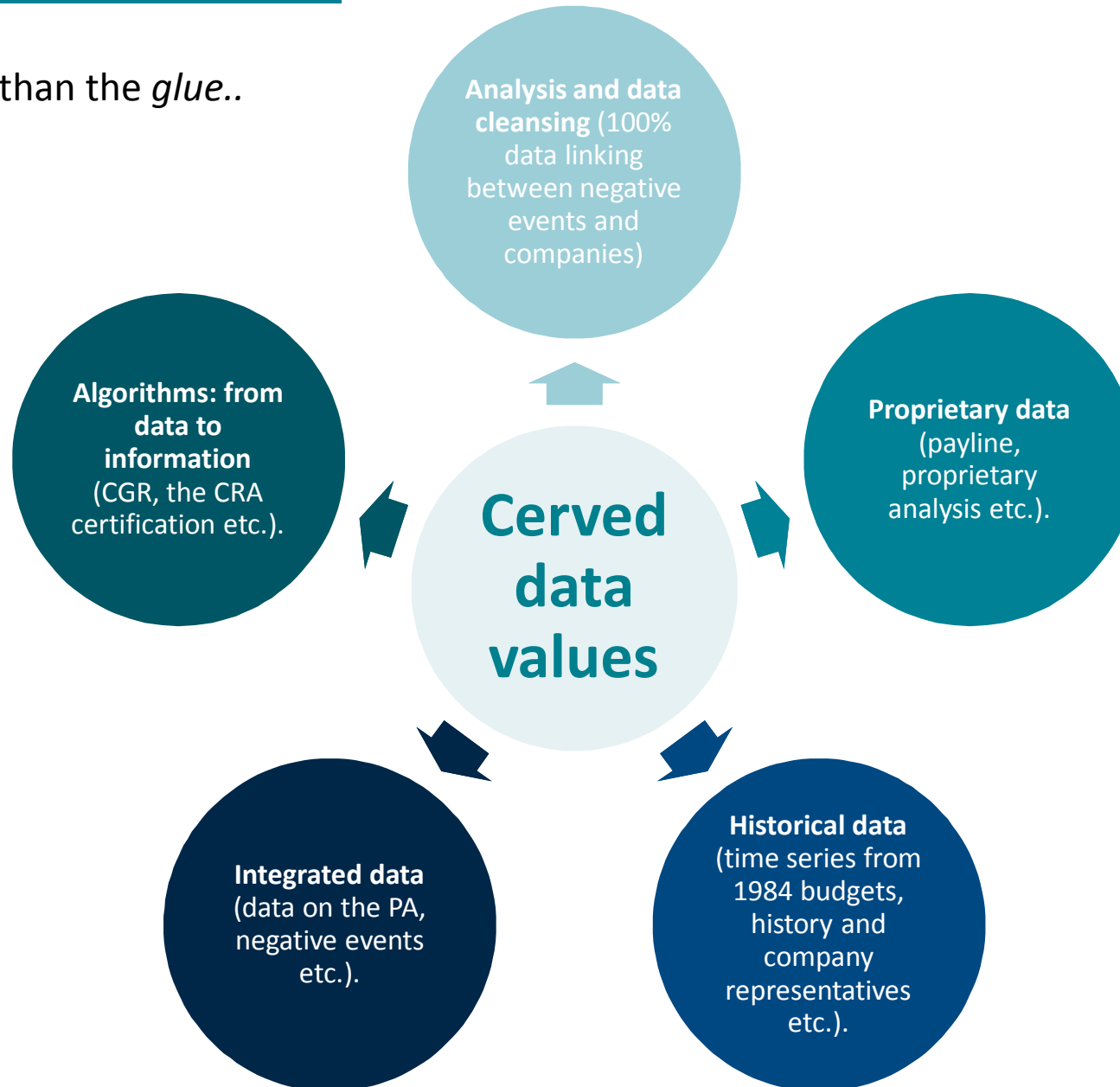
Cerved data vision

We are the *glue* between..

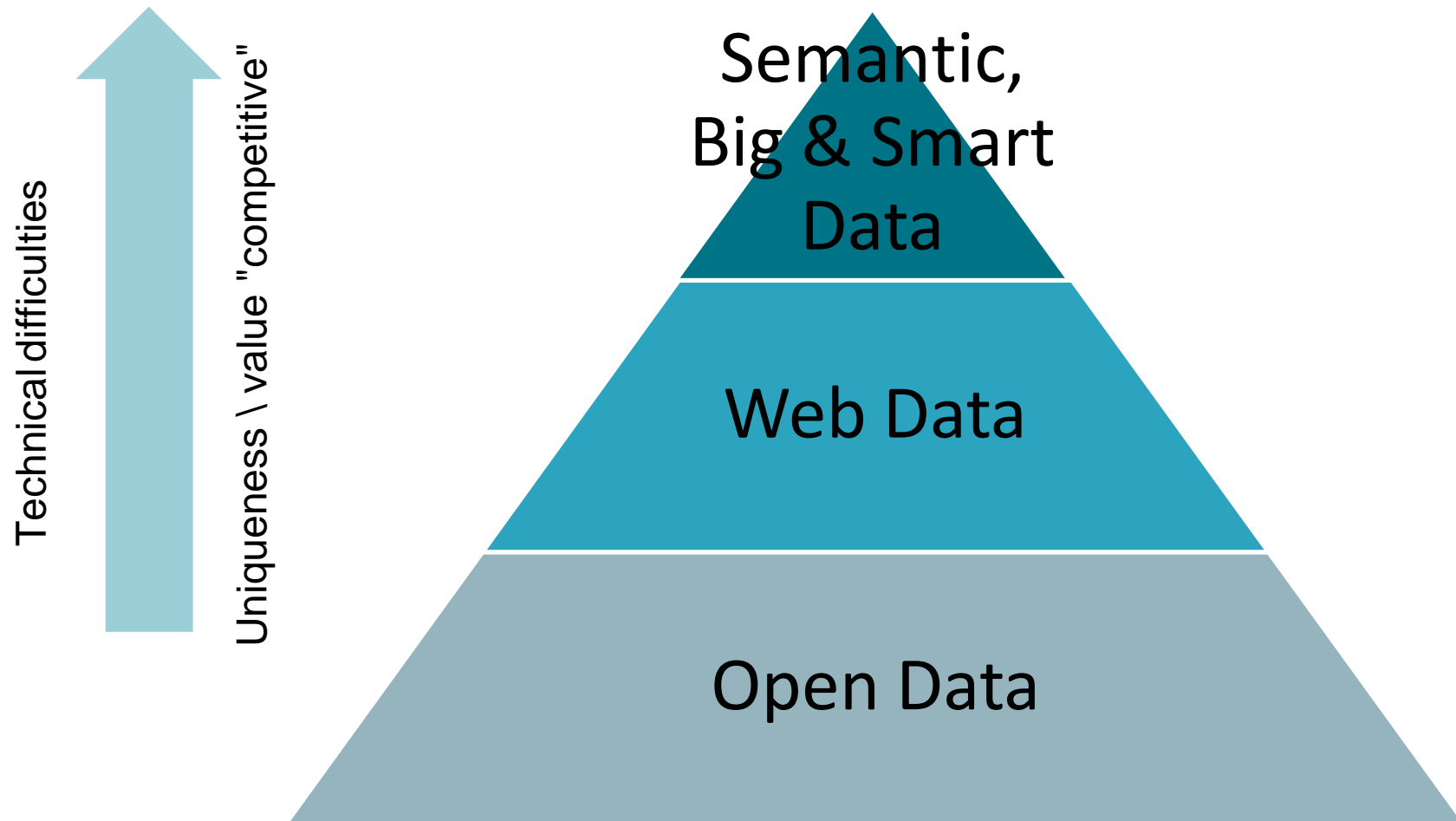


Cerved proprietary data

We are more than the *glue*..



Innovation in data: our pyramid



The top of our pyramid: SpazioDati

INDUSTRY NEWS, SEMANTIC SEARCH

SpazioDati Raises \$1.8M Series A Investment from Cerved Group

By Angela Guess on April 4, 2014 2:30 PM



1.3 milioni da Cerved Group a SpazioDati. Gatti: "Vogliamo farli crescere"

E' un investimento industriale quello operato da Cerved Group ai benefici della startup SpazioDati. Michele Barbera, CEO: "Ci hanno trovati loro. Adesso amplieremo il team e miglioreremo la tecnologia"

Abbonati subito!  Milano 14° (cambia)

24 ORE
NOVA24 TECH
Venerdì • 25 Aprile 2014 • Aggiornato alle 09:23

NEW! Quotidiano del Fisco English version
Professioni e Imprese24 Formazione Eventi
Banche dati Servizi
Versione digitale

HOME ITALIA & MONDO NORME & TRIBUTI FINANZA & MERCATI IMPRESA & TERRITORI **NOVA24 TECH** PLUS24 RISPARMIO COMMENTI & INCHIESTE STRUMENTI DI LAVORO STORE24 Acquista & abbonati

Gadget Social Media Business **Startup** Innovazione Scienza Games App&Entertainment Prove Nòva

Cerved Group investe 1,3 milioni di euro in SpazioDati, startup specializzata in Big Data

di [Gianni Rusconi](#) 1 aprile 2014

IN QUESTO ARTICOLO

Argomenti: Investimenti delle imprese | Cerved | Trento Rise | Pietro Masera | Marco Nespolo | Camera di Commercio | Michele Barbera | Gabriele Antonelli | Pisa



<http://dandelion.eu/>

Free Unlimited plan for Bigdivers! (write us)

dataTXT find places, persons, brands, and events in documents and social media.

dataTXT Features NEX - Named Entity eXtraction SIM - text SIMilarity CL - text Classifier

Enter a sentence to extract named entities. dataTXT-NEX works well also on short texts.

Insert a Text or a URL of a newspaper/blog to analyze with dataTXT-NEX:

The Mona Lisa is a 16th century oil painting created by Leonardo. It's held at the Louvre in Paris.

Try this examples:

Text URL

More keyword More precision

Language: Autodetected Deep Analysis Parse Hashtag

Extract Entities

Hey, did you:
• get a free API key
• get started with dataTXT-NEX
• learn more about dataTXT

Language: English

Show API uri Show me the response

The **Mona Lisa** is a 16th century **oil painting** created by **Leonardo**. It's held at the **Louvre** in **Paris**.

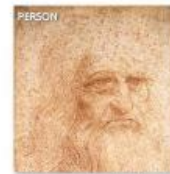
1 person 1 work 0 organisations 2 places 0 events 1 concept



Mona Lisa



Oil painting



Leonardo da Vinci

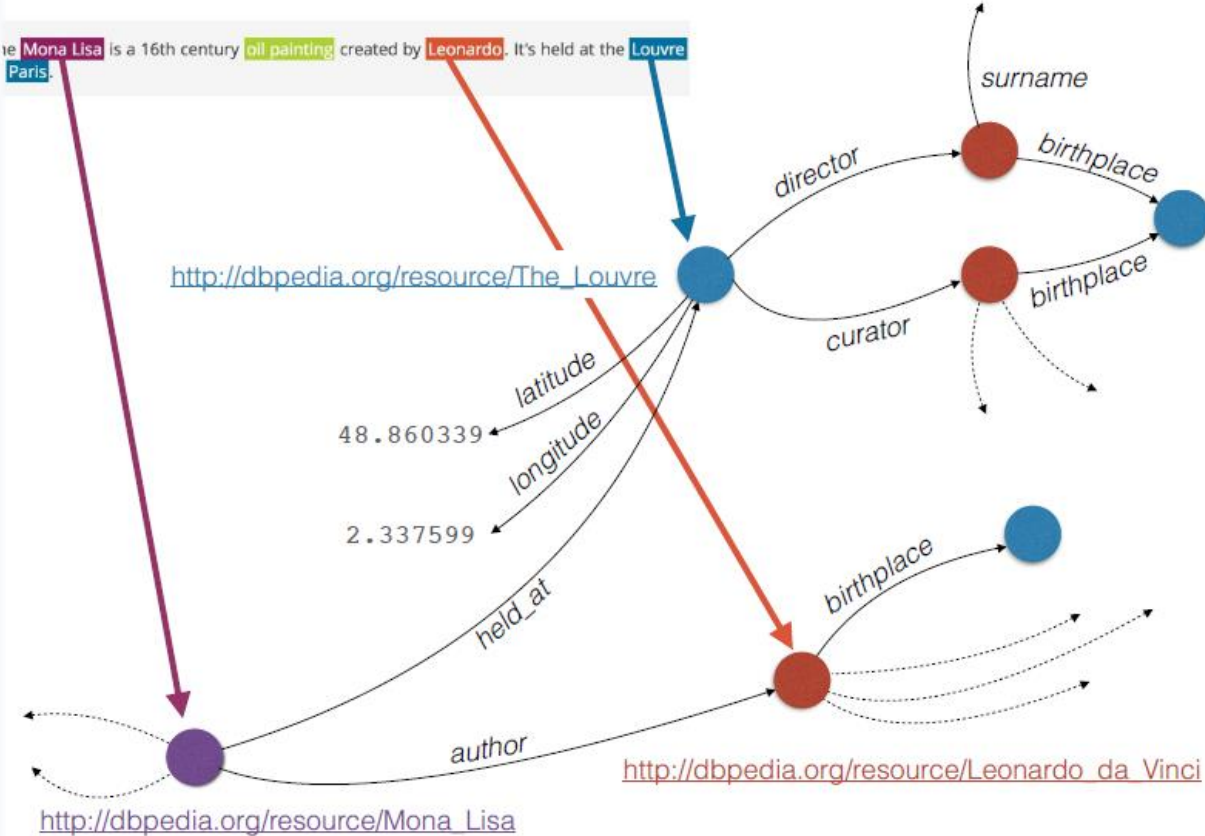


The Louvre



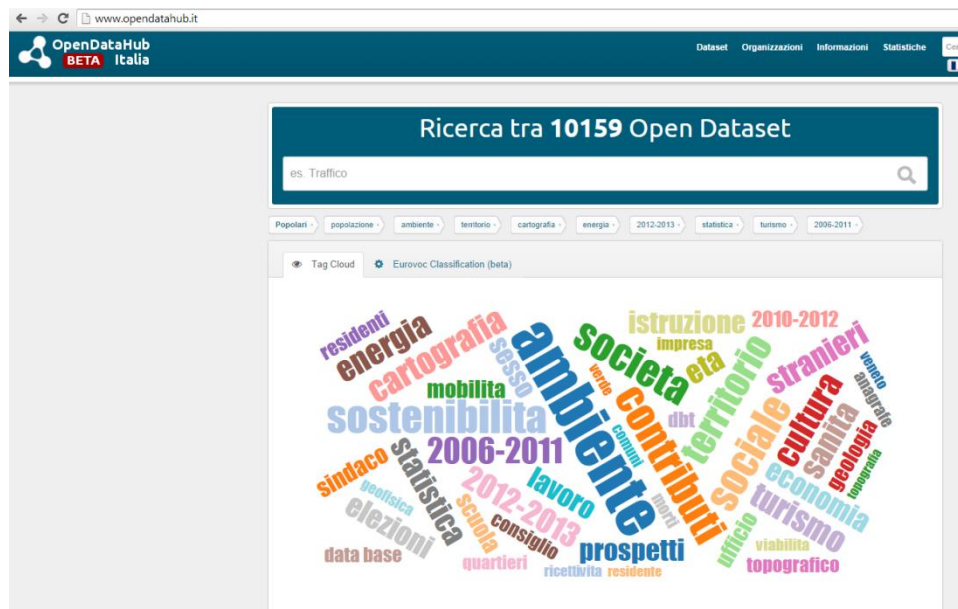
SPAZIODATI

An entry point to the Web of Data



Open Data: Cerved vision - opportunity

Many data from real world ð



proprietary data + open data = big value

McKinsey&Company

McKinsey Global Institute
McKinsey Center for Government
McKinsey Business Technology Office

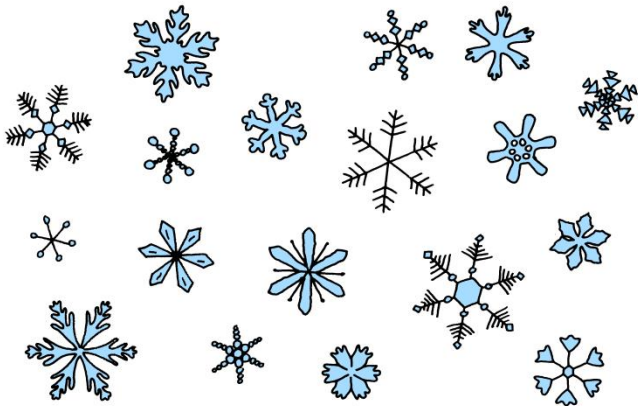


October 2013

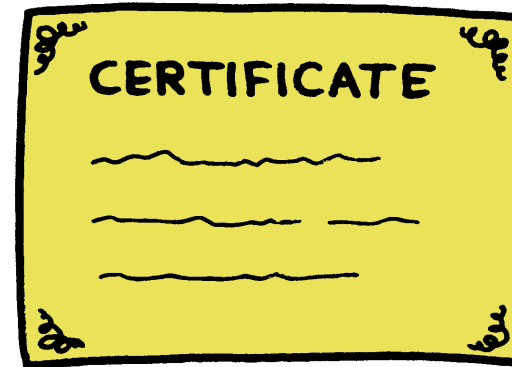
Open data: Unlocking innovation and performance with liquid information

Open Data: Cerved vision - issue

Too different formats



Authoritative source



Update frequency



Quality data problems



Open Data: Tools to accelerate ...

“ Data Management System:

- Document DB (es: MongoDB)
- Graph DB (es: Neo4J)
- RDMS (es: Oracle)

“ Integration tool (es: Pentaho, Open Refine)

“ Data-analysis tool & framework (es: Excel, Refine, Teradata, R, Python)

“ Analytics tools (es: Splunk, Tableau)

“ Agile datascience: WIP

Open Data: Cerved use cases - live

<http://www.pa.cerved.com/portalePA/>

[Home](#) [FAQ](#)

Cerved pubblica
amministrazione **SERVIZI PER E SULLA PUBBLICA AMMINISTRAZIONE**

Servizi informativi, collegamenti, approfondimenti e valutazioni su tutti i soggetti nel perimetro
Pubblica Amministrazione:



Cosa puoi fare

- Valuta la gestione dell'Ente Pubblico e delle sue partecipate
- Identifica le aree di miglioramento
- Confronta il tuo Comune con altri
- Trova le società partecipate della Pubblica Amministrazione
- Conosci i collegamenti tra diversi soggetti
- Poni sotto osservazione un Fornitore o una Partecipata
- Verifica gli incarichi di un Esponente politico
- Richiedi approfondimenti su misura (analisi delle aziende sul territorio di un Ente, analisi del settore e della concorrenza delle Partecipate e dei Fornitori, altro...)

A cosa serve

- **Informati** e conosci tutte le informazioni sui soggetti di tuo interesse, grazie a un bacino informativo unico ed esclusivo
- **Confronta** la tua realtà con i soggetti a te simili
- **Monitora i soggetti** ricevendo notifiche puntuali degli eventi di interesse sui soggetti selezionati

AREA MULTIMEDIALE
Scopri le potenzialità di Cerved PA
attraverso video, presentazioni, brochure...

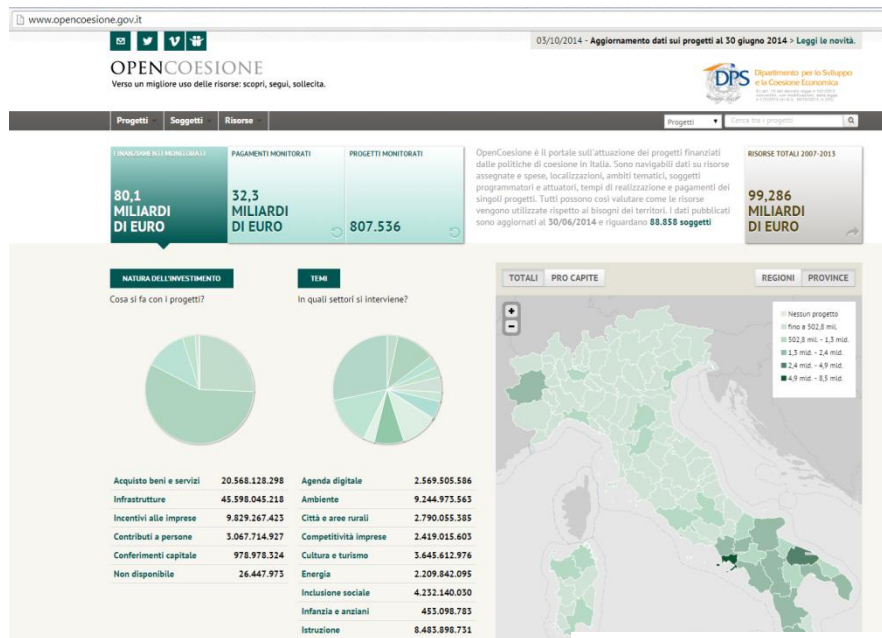
CLICCA
QUI

Cerved PA è parte dell'offerta Cerved Group, leader in Italia nel campo della business information. Le sinergie fra le diverse realtà del gruppo garantiscono informazioni attendibili, sempre aggiornate, puntuali, tempestive.

Tweets

[Follow](#)

Open Data: Cerved use cases - wip



Fondi Strutturati

Open data by [OpenCoesione](http://www.opencoesione.gov.it), portale sull'attuazione degli investimenti...

| Progetto | Valore | Pagamenti | Stato | Informazioni |
|--|--------|---|----------|---|
| Introduzione Di Strumenti E Metodologie Per L'Innovazione E Lo Sviluppo Organizzativo Dell'Impresa | 49.976 | 80% visualizza pagamenti | in corso | INIZIO (PREVISTO): 31.10.11 (EFFETTI): 30.11.11 FINE (PREVISTA): 31.10.14 (EFFETTI): n.d. TEMA: Acquisto servizi reali PROGRAMMATORE: Regione emilia romagna Altre informazioni |
| Gestione Dei Processi E Sviluppo Delle Performance Aziendali In Pasquali Srl | 38.878 | 100% visualizza pagamenti | chiuso | INIZIO (PREVISTO): 31.10.11 (EFFETTI): 30.11.11 FINE (PREVISTA): 31.10.13 (EFFETTI): 31.10.13 TEMA: Acquisto servizi reali PROGRAMMATORE: Regione emilia romagna Altre informazioni |

Dati aggiornati al 31/08/2013.

Data Quality: a strategic step in datascience

The cost of data cleansing: an example

03/10/2014 - Aggiornamento dati sui progetti al 30 giugno 2014 > [Leggi le novità.](#)

OPENCOESIONE
Verso un migliore uso delle risorse: scopri, segui, sollecita.

DPS Dipartimento per lo Sviluppo e la Coesione Economica
Ex art. 10 del decreto legge n. 101/2013 convertito, con modificazioni, dalla legge n. 130/2013 del G.U. 30/06/2013, n. 255

Progetti Soggetti Risorse

Progetti Cerca tra i progetti

CUP: B37I10000000004
CONTRATTO DI PROGRAMMA MER MEC SPA - MER MEC SPA (CAPOFILA)
Data di aggiornamento: 30.06.2014
INCENTIVI ALLE IMPRESE - ATTIVITA' DI RICERCA
COMPETITIVITÀ IMPRESE - SERVIZI DI SOSTEGNO AVANZATO PER IMPRESE E GRUPPI DI IMPRESE

SEI UN PROTAGONISTA DI QUESTO PROGETTO? [RACCONTACELO QUI.](#)

SOGGETTI

| | |
|---------------------------------|--------------------------|
| PROGRAMMATORE REGIONE PUGLIA | ATTUATORE MER MEC SPA |
|---------------------------------|--------------------------|

TEMPI

| | |
|-----------------------------------|------------------------------------|
| INIZIO PREVISTO 26 giugno 2009 | INIZIO EFFETTIVO 26 giugno 2009 |
| FINE PREVISTA 26 giugno 2009 | FINE EFFETTIVA 26 giugno 2009 |

PRIORITÀ QSN

TERRITORIO MONOPOLI

RISORSE PUBBLICHE

| | |
|------------------------------------|---|
| FINANZIAMENTO 3.000.000,00 euro | PAGAMENTI EFFETTUATI 2.700.000,00 euro |
|------------------------------------|---|

90%

| | |
|---|----------------|
| Unione europea | 1.749.000 euro |
| Fondo di Rotazione (Co-finanziamento nazionale) | 726.300 euro |
| Regione | 524.700 euro |

FONDO EUROPEO DI SVILUPPO REGIONALE (FESR)
Fondi Strutturali relativi alla programmazione 2007/2013
PROGRAMMA
POR CONV FESR PUGLIA
ASSE
RICERCA E INNOVAZIONE PER LA COMPETITIVITA'
OBIETTIVO
1a) Elevare la domanda e la propensione delle imprese ad investire in Ricerca

Data Quality: a strategic step in datascience

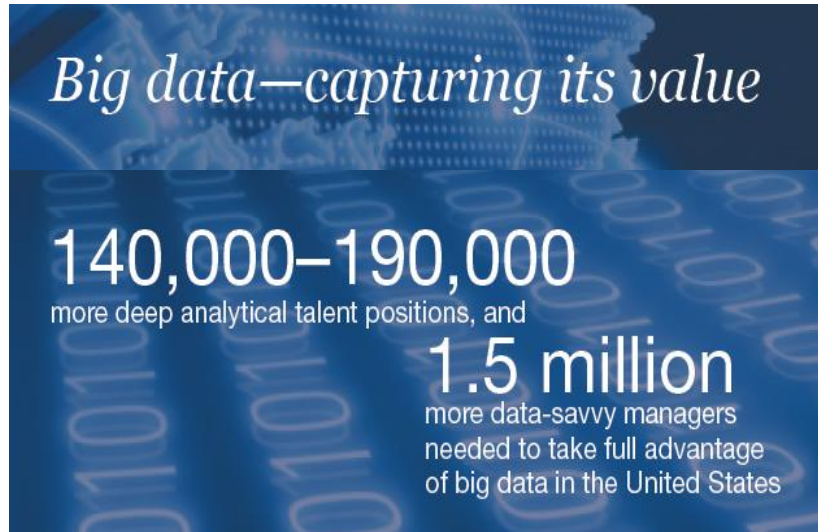
The cost of data integration: an example

| FORNITURA | TOTALE | Distinti CERTI | DUBBI | Creati |
|--------------------------------------|--------|----------------|--------|--------|
| <u>Opencoessione</u> - Attuatore | 74.987 | 48.118 | 26.327 | 0 |
| <u>Opencoessione</u> - Programmatore | 3.364 | 2.896 | 439 | 0 |

34% senza matching certo!

Some (not definitive) thoughts about datascience

McKinsey : an optimistic view?



My optimistic view ð .

“... Software is eating the world ...”

2011 - MARC ANDREESSEN



“ ... together with Data ... ”

2014 – Common Sense

Some (not definitive) thoughts about datascience

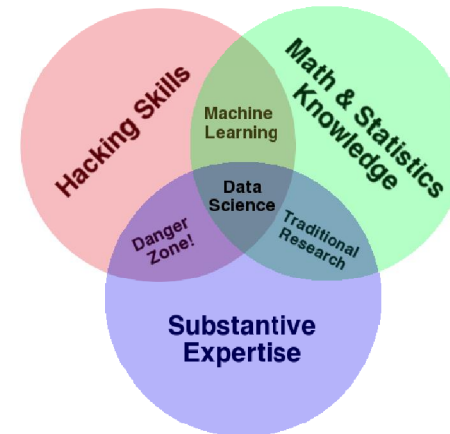
The future of Data Science is smarter tools, not smarter humans. Really?

The screenshot shows a blog post from Oracle's 'The Data Warehouse Insider' series. The title is 'The End of the Data Scientist Bubble...' by Jean-Pierre Djikj, dated September 29, 2014. The article's main argument is that the current hype around data scientists is a bubble that is about to burst. It suggests that the future of data science lies in smarter tools rather than smarter humans. The article is categorized under 'The Demise of the Human Data Scientist'.

Not all people think like Oracle ð

UNIVERSITY of WASHINGTON

Drew Conway's Data Science Venn Diagram



4/28/13

Bill Howe, UW

3

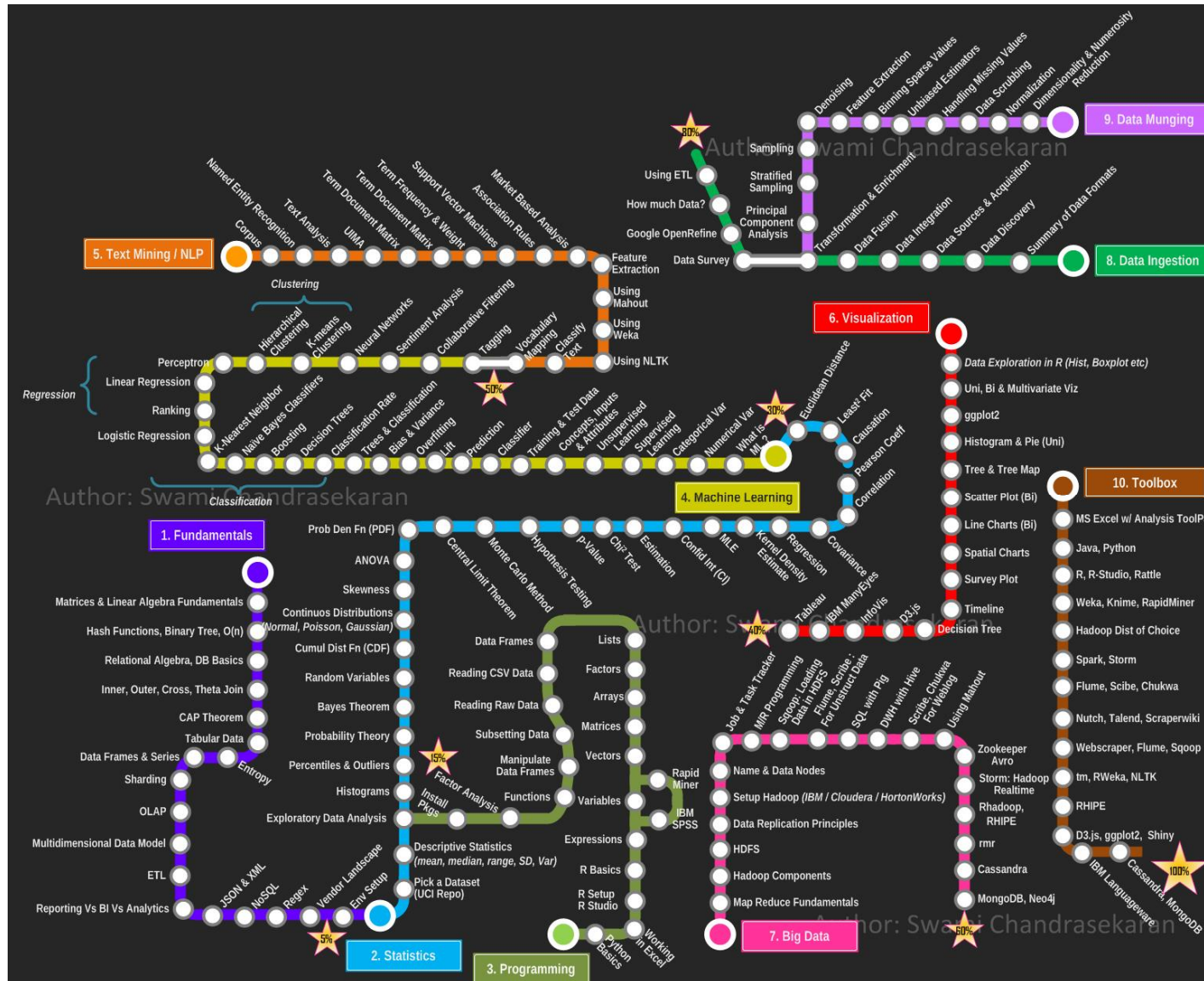
The screenshot shows a blog post titled 'The end of the Data Scientist Bubble' by Mirko Krivanek, dated October 1, 2014 at 8:00am. The post discusses a provocative article from Oracle's blog and argues that many companies are confusing the title of 'data scientist' with the actual skills required for the job. It suggests that many people who call themselves data scientists are not truly data scientists and are not leveraging their success stories. The post includes a photo of a man in a suit pulling open his shirt to reveal a 'DS' logo on his chest, symbolizing the 'Data Scientist' identity.

Fonti: <http://drewconway.com/>

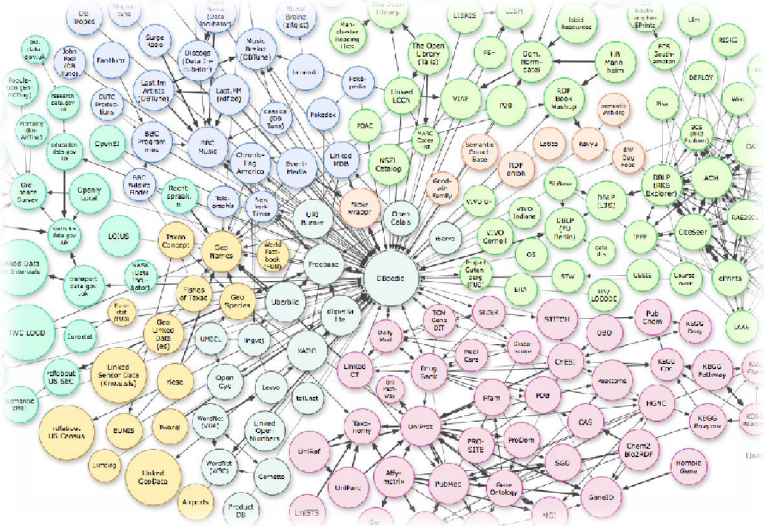
https://blogs.oracle.com/datawarehousing/entry/why_the_data_scientist_bubble

<http://www.datasciencecentral.com/profiles/blogs/the-data-scientist-buble-has-started-to-explode>

Never ending travel...



“Il futuro non è più quello di una volta...”



Now & tomorrow ã

