CRISP centro di ricerca interuniversitario per i servizi di pubblica utilità

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

How **Big Data Analytics** can Improve Your Business

Innovations in Data Science:
Insights Platform

Hints from a real case

Real Time Labour Market Intelligence

on skill requirements

Emsi | burning glass

Mauro Pelucchi is a senior data scientist and big data engineer
responsible for the design of the "Real-Time Labour Market Information System on Skill Requirements" for CEDEFOP.

He currently works for EMSIBG as Head of Global Data Science Team.
His main tasks are related to machine learning modelling, labour market analyses, the design of big data pipelines to process large datasets of online job vacancies and the release of new innovative solutions about LMI.

In collaboration with the University of Milano-Bicocca, He took part in many research projects related to the labour market intelligence systems.

He collaborates with the University of Milano-Bicocca as a lecturer at the Master Business Intelligence and Big Data Analytics and with the University of Bergamo as a lecturer in Computer Engineering.

Emsi | burning glass

https://github.com/mauropelucchi/sgi_workshop_2022

Emsi | burning glass

Agenda

1. Using online Job Vacancies to create Labour Market Intelligence
2. A successful Big data project
3. Traditional vs Big Data: How build an Insight Platform
4. Workshop @ Databricks
5. Validation Exercise
6. Data analysis path
7. Design, implementation and evolution

Emsi | burning glass

Using online Job Vacancies to create Labour Market Intelligence

Skills OVATE

Online Vacancy Analysis Tool for Europe

Data classification

Merging different vacancy templates, styles and vocabularies into single language, using European and global taxonomies

Occupation:
up to ISCO 4-digit
400+ occupations

Skill:
ESCO version 1
2 000+ skills

Sector:
NACE rev. 2
20+ sectors

Region:
NUTS-2
276 regions

Seven countries
Czechia, France, Germany, Italy, Ireland, Spain & United Kingdom

Top skills requested

Cross-cutting skills

Importance of **data analysis skill** across jobs

Jobs demand

Stay tuned for more!

28 countries covered by end 2019
Fully operational system in 2020
Collaboration with ESCO & ESTAT

rlmi@cedefop.europa.eu
www.cedefop.europa.eu/skills-online-vacancies

(end March 2019)

Covering periods
1 July to 31 December 2018

Let me to show somethings

- Skillspanorama – Skills in Online Vacancies
  - https://skillspanorama.cedefop.europa.eu/en/indicators/skills-online-vacancies
- Skills OVATE
  - https://www.cedefop.europa.eu/en/data-visualisations/skills-online-vacancies
- EMR Notebook and Google CoLab
  - Jupyter

Emsi | burning glass

Revealed comparative advantage

LAB SESSION

RCA

Given a set of occupations $\bar{O} = \{o_k, k = 1, \ldots, m\}$, a set of skills $\bar{S} = \{s_j, j = 1, \ldots, p\}$, and a matrix $\underset{m \times p}{M}$ that contains a value of sf for each pair of occupations $o_k \in \bar{O}$ and skills $s_j \in \bar{S}$, rca for o_i and s_l is defined as:

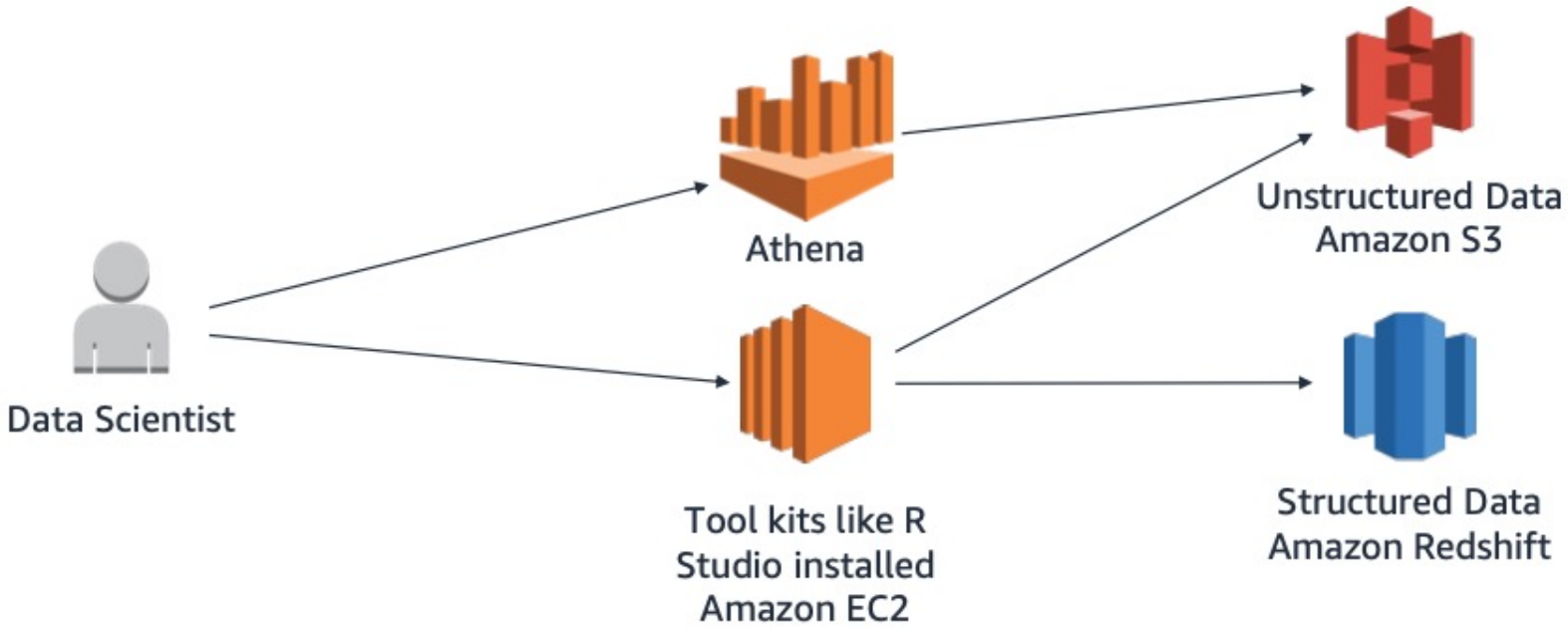
$$rca(o_i, s_l) = \frac{sf(o_i, s_l) / \sum_{j=1}^{p} sf(o_i, s_j)}{\sum_{k=1}^{m} sf(o_k, s_l) / \sum_{k=1}^{m} \sum_{j=1}^{p} sf(o_k, s_j)}$$

Emsi | burning glass

Occupation similarity

From the notebook:

- We will select a dataset
- We will calculate the RCA
- And we will see our create bridge between occupations



Emsi | burning glass

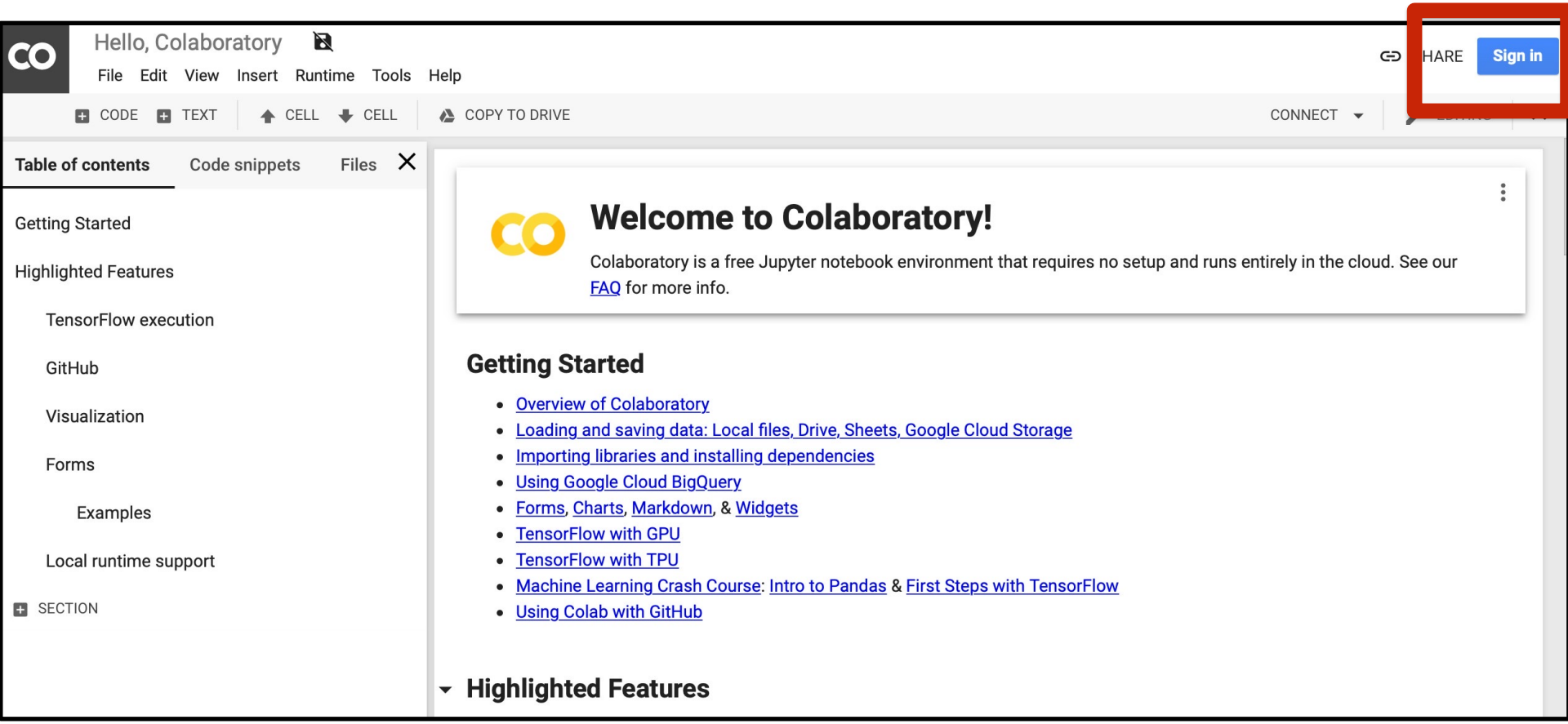
What is Google Colaboratory?

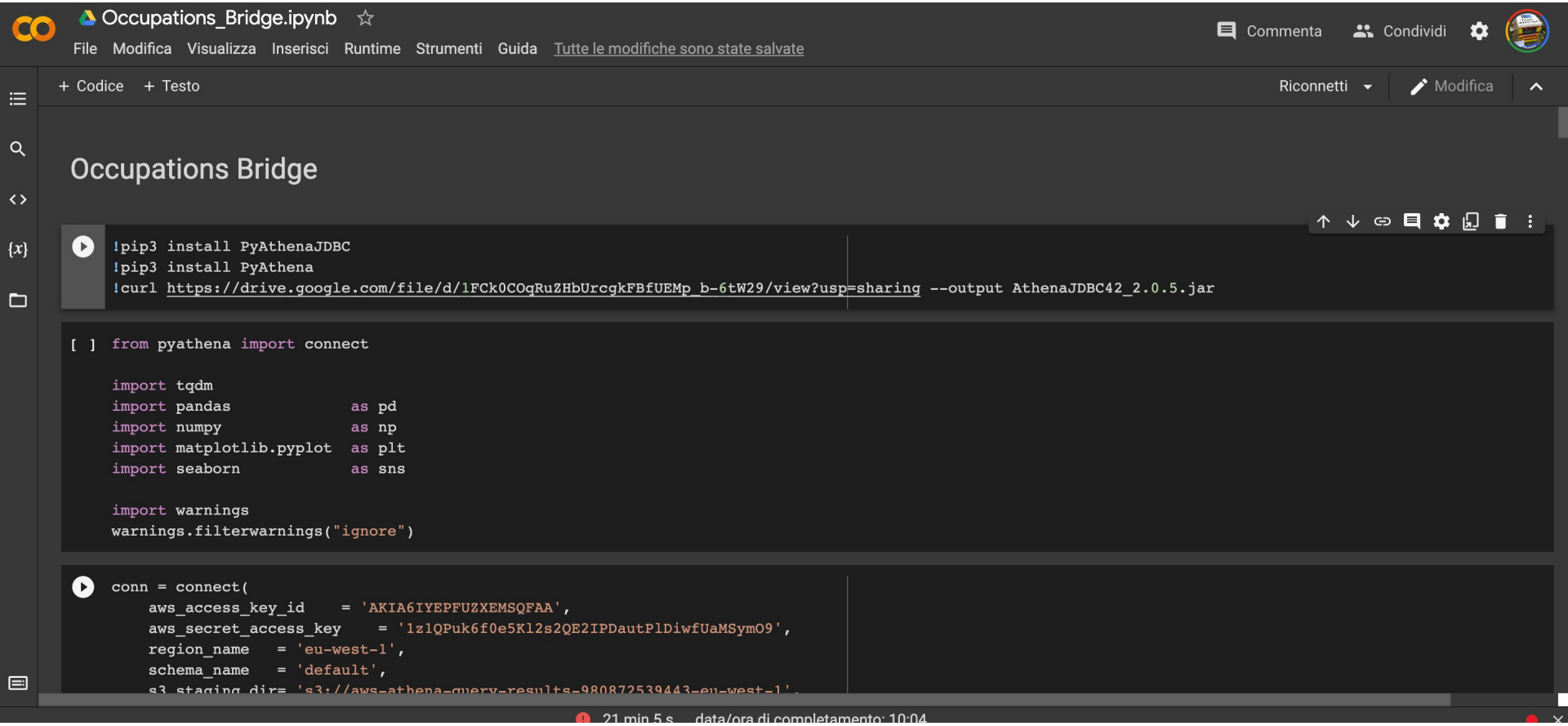
- Free Jupyter notebook environment
  - About 13G memory and 25G disk space limit
- No setup
- 100% Cloud
- Python + R
- Hardware accelerator
  - Graphical Processing Unit (GPU)
  - Tensor Processing Unit (TPU)

Awesome Sharing Feature

- Save Colab to Google Drive and share.
- Save Colab to GitHub directly.
- Upload / Download Colab to local drive.
- Share with others users
- Control revisions
- Google document
- Install packages

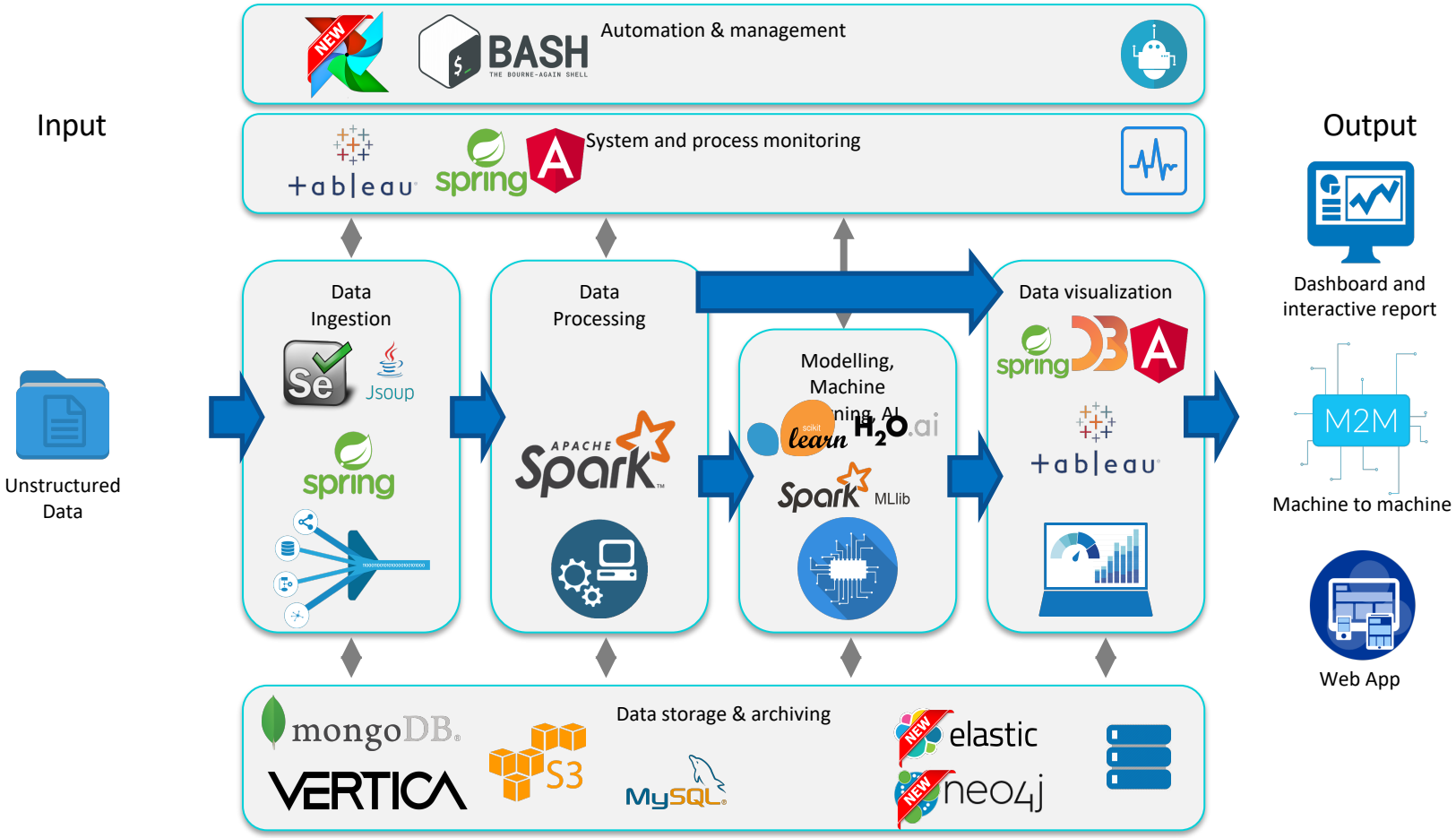
https://colab.research.google.com/notebooks/welcome.ipynb





Emsi | burning glass

Organic view



Emsi | burning glass

A SUCCESSFUL BIG DATA PROJECT:

LABOUR MARKET INFORMATION AND INTELLIGENCE

Emsi | burning glass

Labour Market Information and Intelligence

LMI is any quantitative or qualitative facts, analysis or interpretation about the past, present or future structure and workings of the labour market and the factors that influence it.

Economic and labour market conditions

Education, qualifications, training and skills

Current and future demand and supply of labour and jobs

Vacancies and recruitment

Emsi | burning glass

Needs: new tools for LMI

- Famous study of **Frey and Osborne** (THE FUTURE OF EMPLOYMENT, Oxford)
  - 47% of Jobs will disappear in the next 25 years.
- 65% of children entering primary school today (2017) will ultimately end up working in completely new job types that don't yet exist.
- Huge implications in terms of skill requirements
  - Numbers are worrying but are they really true?
- We need to implement several complementary tools for investigating these changes

Emsi | burning glass

Why Big Data Analytics for LMI?

- Lacking data on skill demands by employers
- Conventional methods are
  - expensive
  - suffer from time lags
  - focus on specific types of skills
  - Surveys are rigid and lengthy tools
- Forecasting tools to identify the most relevant trends
  - But forecasting tools are necessarily imprecise about the features and skill requirements of the jobs of the future
- Skills anticipation
- Useful
  - Understand the real market demands
  - Inform career mobility and training choices
  - Fine-tune training offer

Emsi | burning glass

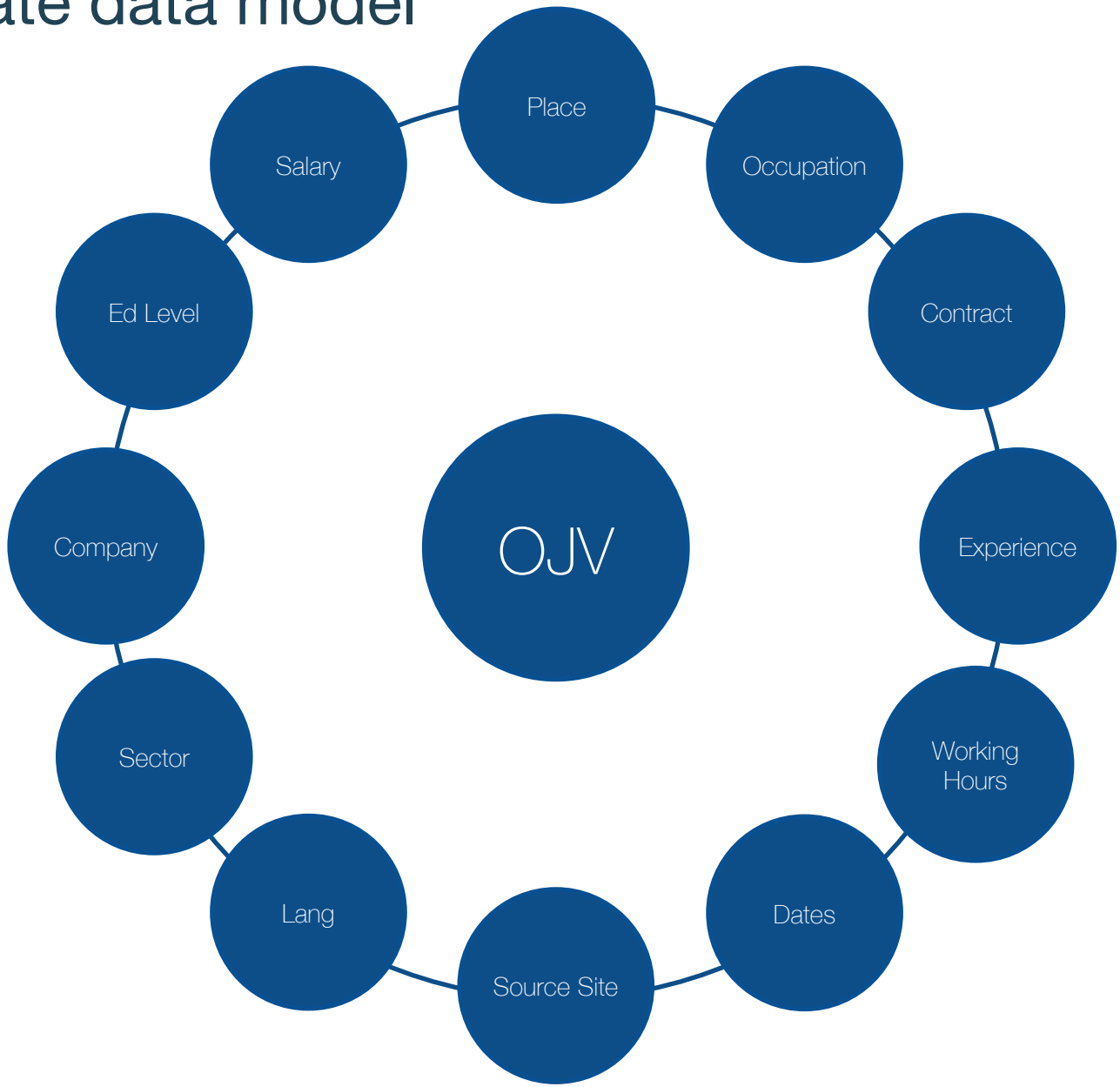
Traditional vs Big Data

DESIGN AND BUILD A SYSTEM OF INSIGHTS FROM BIG DATA

Challenges

- Handle a huge **amount** of near real time data
- Data coming from web → Need to detect and reduce **noise**
- **Multi language** environment
- Need to relate to **classification standards**
- Find a way to **summarize and present** a wide and complex scenario

Aggregate data model



Major steps

1. Landscaping
    - Context, specific reports, source identification
2. Early release
    - Preliminary results, only 6 countries (60% of EU population) to get a preliminary idea of what output the project can produce
3. Final version
    - All EU countries
    - Freely available (open data and methods)

Emsi | burning glass

Landscaping

~30 international experts

How important are OJVs as a recruitment method in the EU?

What are the differences in coverage within sectors, occupations, education levels and regions?

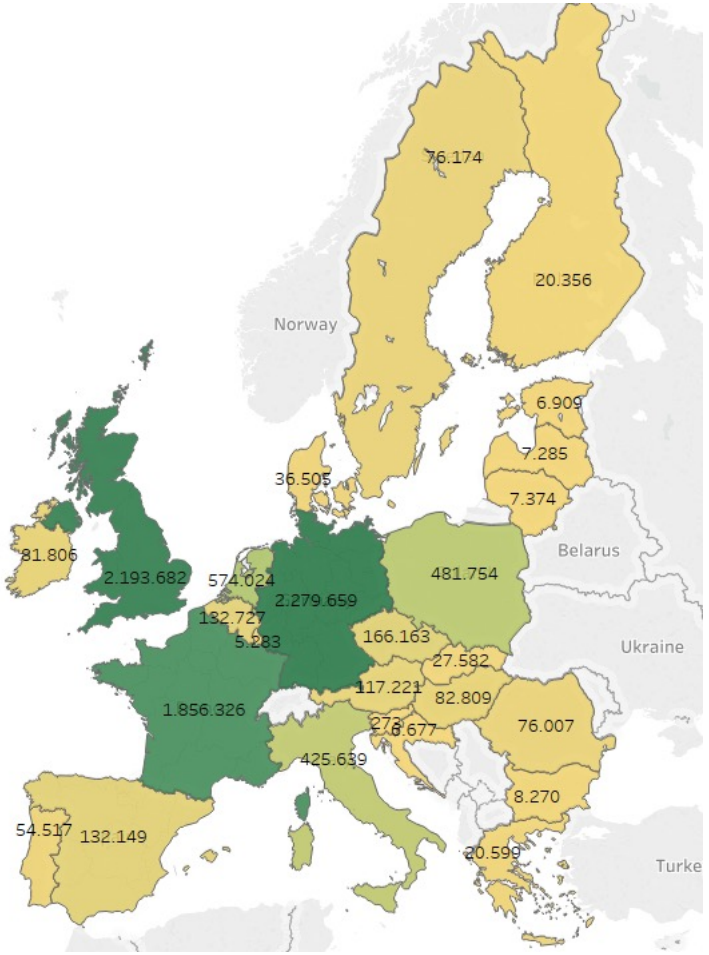
Which are the most important online job-portals in the EU?

What information do OJVs contain?

Sources identification and selection

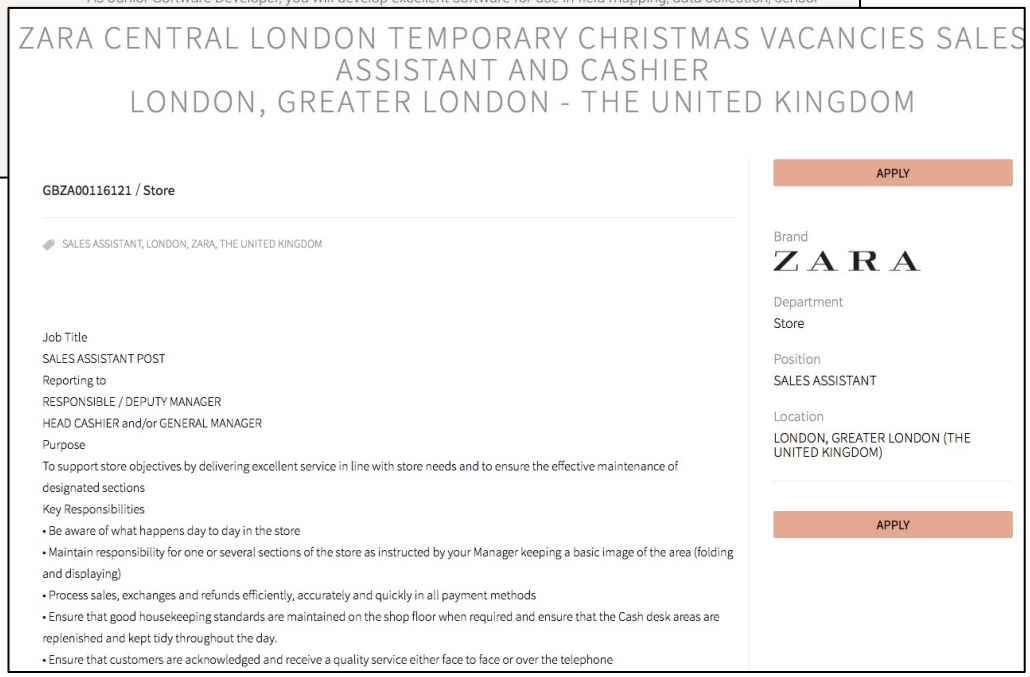
High number of potential sources

- **530** sources from landscaping
- **109** transnational sources (**20** distinct sources)
- **364** local sources
- Over **40 millions** estimated OJV (high redundancy)
- 25 official languages + 5 semi official languages



Number of sources for online job vacancies

Application of Big Data Analytics
Real Time Labour Market Intelligence



Noise

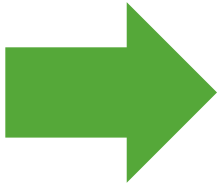
Coverage among all the countries and occupations

No standard

Unstructured data

Without a high-quality classification system, there will be no high-quality big data analysis.

Job ads



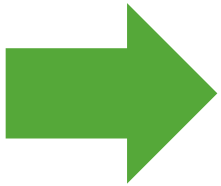
Information Extraction

- Occupation
- Skills
- Time
- Area
- Industry
- …

Junior Software Developer

As Junior Software Developer, you will develop excellent software for use in field mapping, data collection, sensor networks, street navigation, and more. You will collaborate with other programmers and developers to autonomously design and implement high-quality web-based applications, restful API's, and third party integration.

We're looking for a passionate, committed developer that is able to solve and articulate complex problems with application design, development and user experiences. The position is based in our offices in Harwell, United Kingdom.

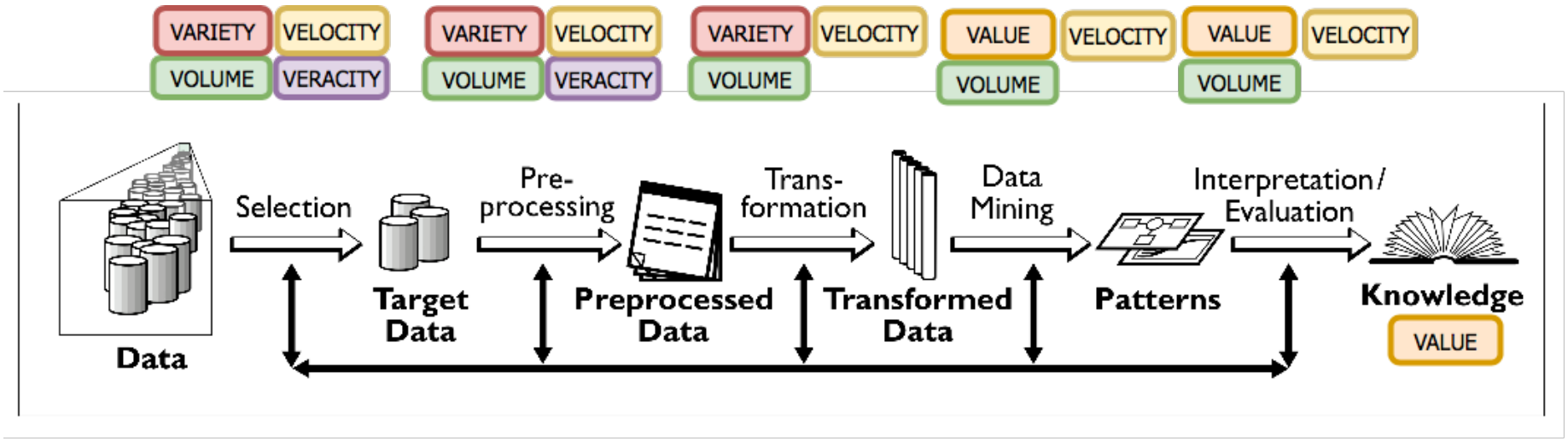


Information Extraction

- 2512 – Software Developer
- Skills: develop software, implement web based applications, problem solving, develop user experiences
- Harwell, UK
- …

Methodological background

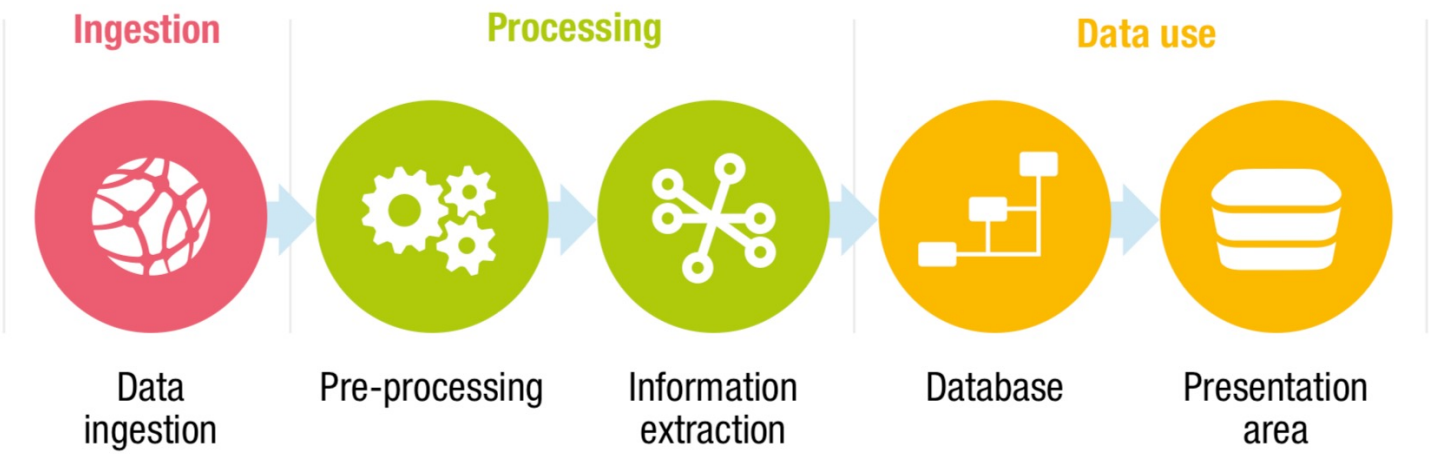
KDD – Fayyad, 1997



Emsi | burning glass

Our approach

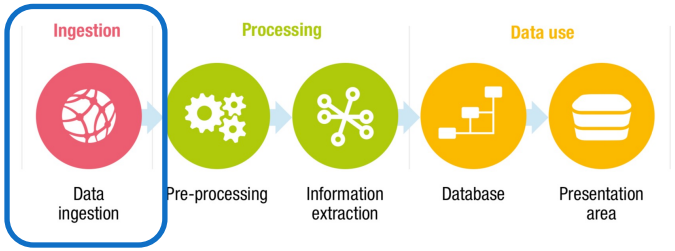
KDD 4 LMI



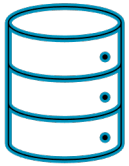
Let's take a deeper look on this framework

Emsi | burning glass

Data Ingestion



The process of obtaining and importing data from web portals and storing them in a Database



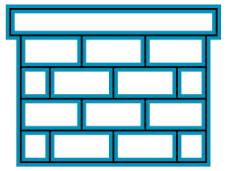
Focus on volume

Coverage augmentation

Balance between quality and effort

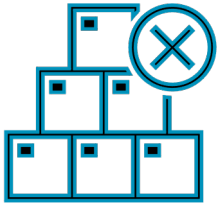
Emsi | burning glass

Data Ingestion - Goals



Robustness
of process

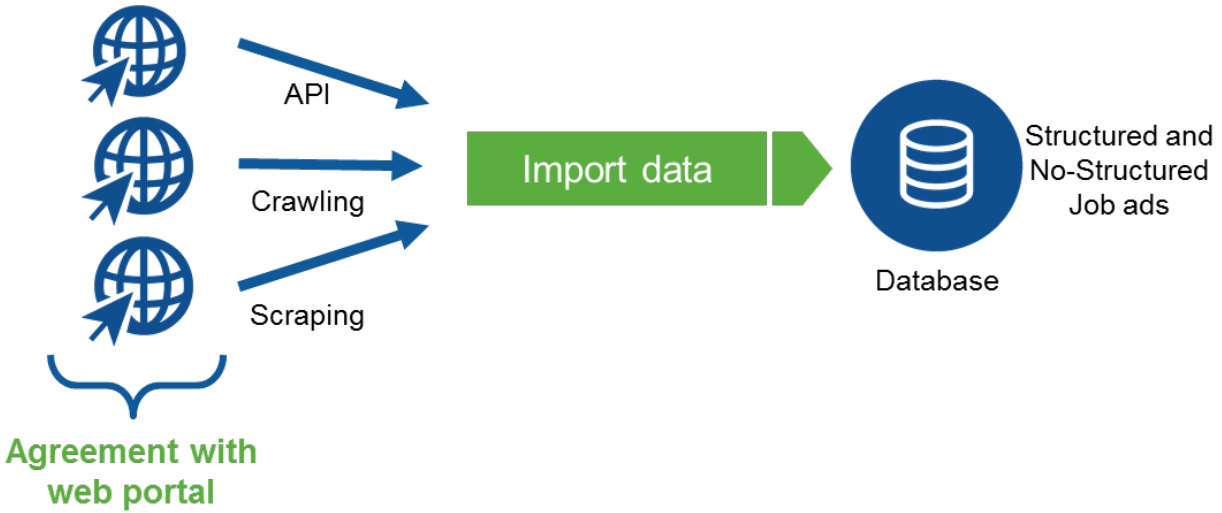
Quality
of data collected



Scalability and
governance

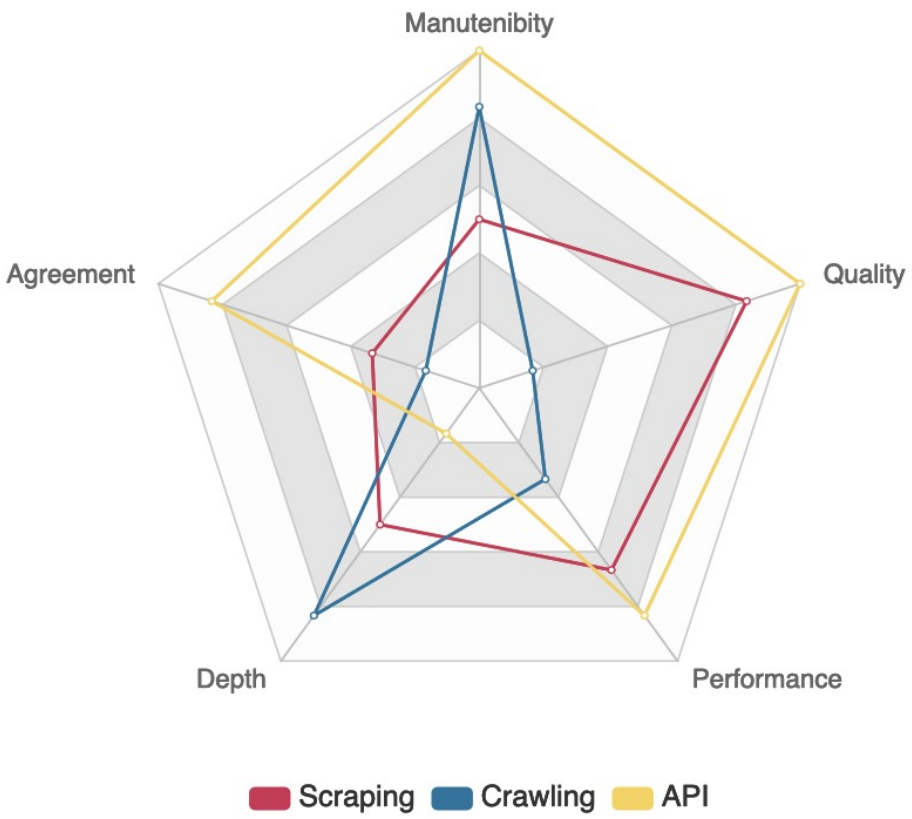
Emsi | burning glass

From the Internet to the database



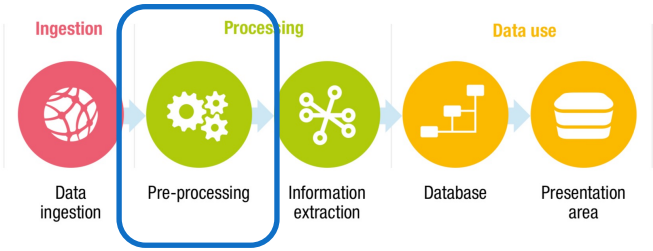
Emsi | burning glass

Data Ingestion



Emsi | burning glass

Data pre-Processing



The process of cleaning ingested data and deduplicating OJVs, to guarantee that analytical phase'll work on data at the highest quality possible

Language detection

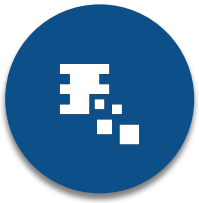
Noise reduction

OJVs Deduplication

Emsi | burning glass

Challenges

How do we measure the quality of collected data?



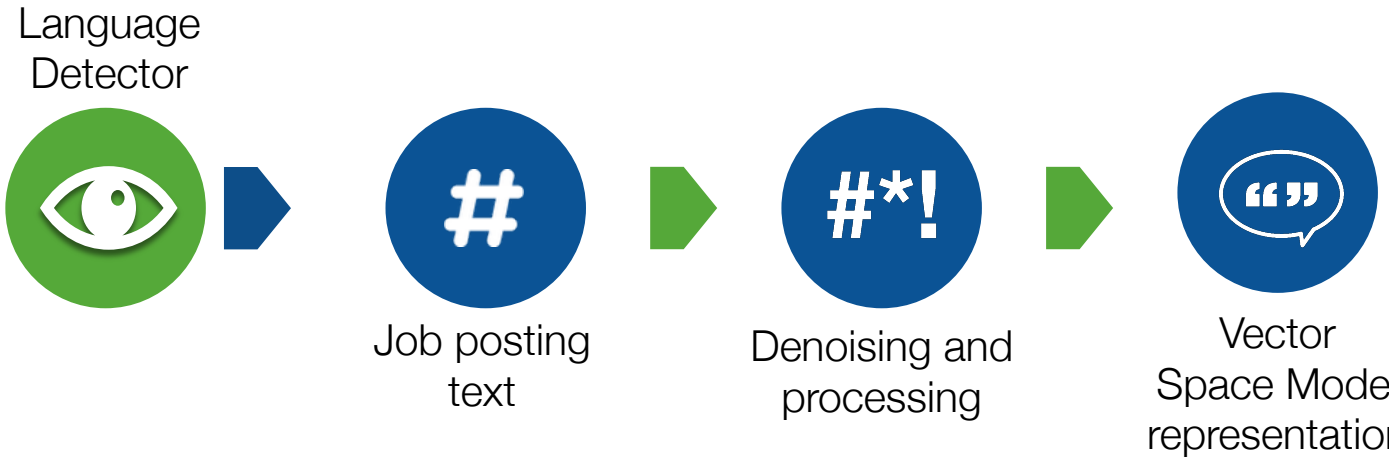
How do we increase the quality of our data?

Which variables affect quality of data?

How do we keep track of quality of data?

Emsi | burning glass

Text pre-processing and summarizing



All the components are developed to work in multi-language environment

For example, stop-words and our ontologies cover about 60 languages

Language Detector (Naïve Bayes classificator trained on Wikipedia corpus) recognizes 60 languages with an accuracy of ~99%

Emsi | burning glass

Noise filter pipeline

Header filter

Looks at page headers (title and other metadata) to judge if forged or not

Content filter

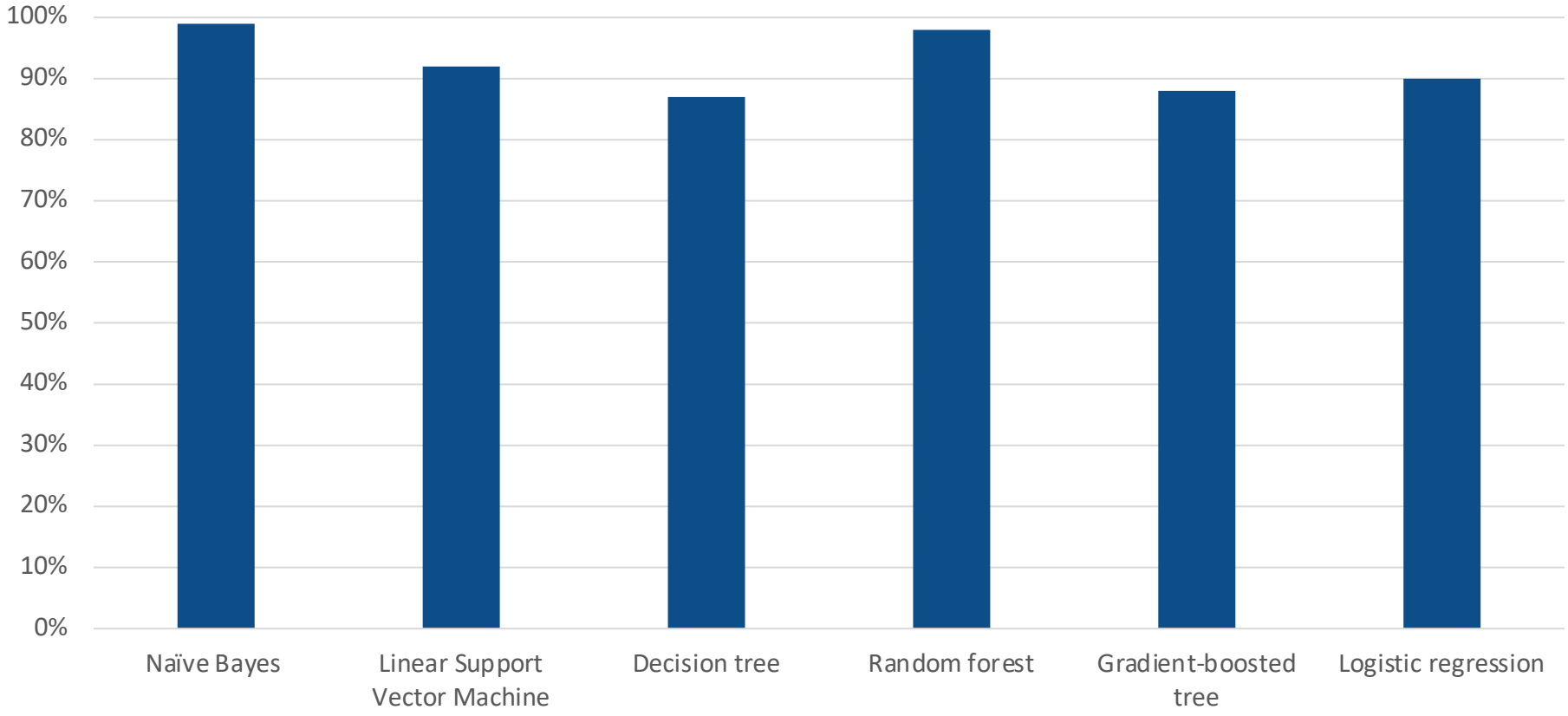
Scans text content of pages with fuzzy logic

It's particularly effective to filter out noise written in several languages.

Machine Learning filter

Emsi | burning glass

Algorithms Used in Noise Detection Which Algorithm is Better?

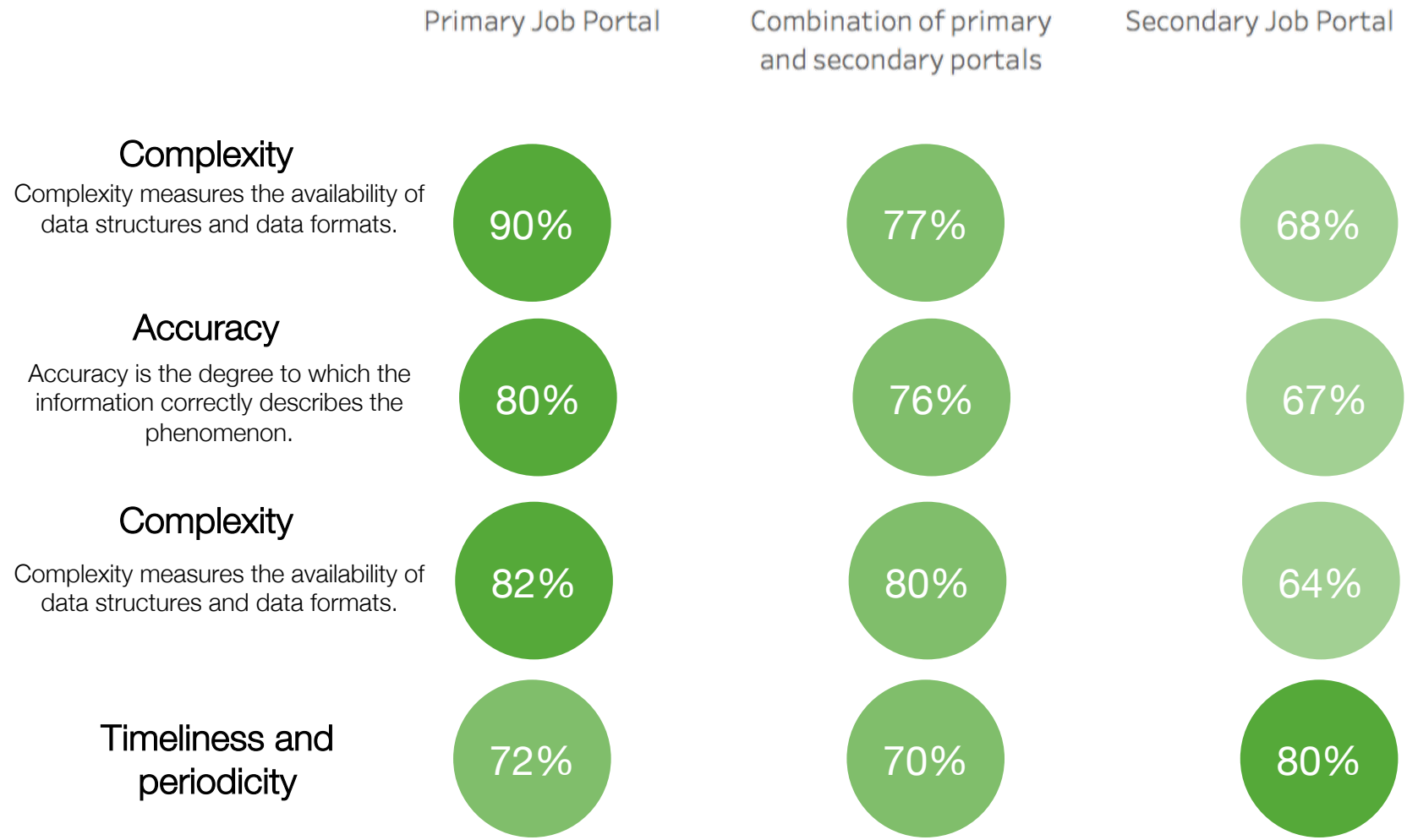


Which Algorithm is Better?

- Train Data: ~14.000 pages; ~50% noise (balanced dataset) for each language
  - Training time: about 30 minutes
- Evaluation Data: ~6.000 new pages for each language
- Naïve Bayes reported to do very well
  - **Results: Job Ads precision 99%, Job Ads recall 97%**
  - More complex algorithms performs similarly, but need more time (!!!) to train

Emsi | burning glass

Measure of quality



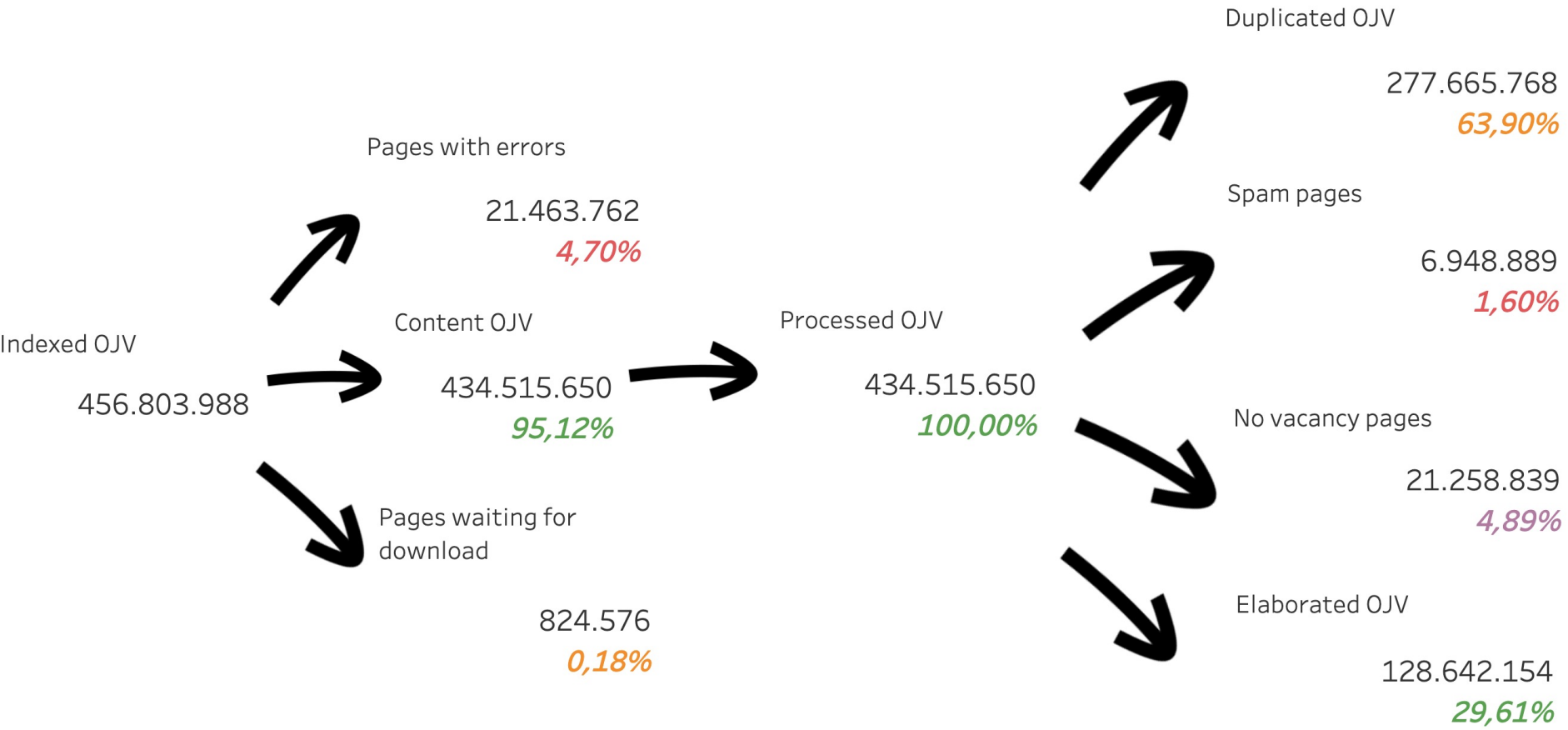
Emsi | burning glass

UNECE Big Data Quality Framework

		Phase of Business Process		
		Input	**Throughput**	**Output**
Hyper-dimension	**Source**	Institutional/Business environment Privacy and Security		Institutional/Business environment Privacy and Security
	Metadata	Complexity Completeness Usability Time-related factors Linkability Coherence / consistency Validity	Processing Principles: 1. System Independence 2. Steady States 3. Application of quality gates	Complexity Accessibility and Clarity Relevance
	Data	Accuracy and Selectivity Linkability Coherence / consistency Validity		Accuracy and Selectivity Validity Coherence / linkability Coherence / consistency Time-related factors

Emsi | bu og

Some numbers – The overall workflow



2018/7 to 2021/6

Content

133.456.787

Processed

127.097.986

Elaborated

76.646.486

Contract
Structured fields collected
(% over total elaborated OJV)

29,90%

Total fields extracted
(% over total elaborated OJV)

76,10%

Method

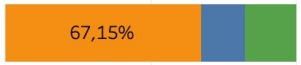
Educational level
Structured fields collected
(% over total elaborated OJV)

8,07%

Total fields extracted
(% over total elaborated OJV)

99,64%

Method



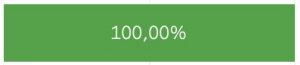
Experience
Structured fields collected
(% over total elaborated OJV)

5,00%

Total fields extracted
(% over total elaborated OJV)

43,91%

Method



Expire date
Structured fields collected
(% over total elaborated OJV)

17,02%

Total fields extracted
(% over total elaborated OJV)

100,00%

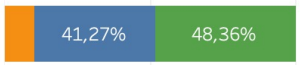
Industry
Structured fields collected
(% over total elaborated OJV)

32,86%

Total fields extracted
(% over total elaborated OJV)

100,00%

Method



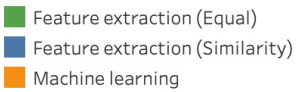
Occupation
Structured fields collected
(% over total elaborated OJV)

7,53%

Total fields extracted
(% over total elaborated OJV)

100,00%

Method



Place
Structured fields collected
(% over total elaborated OJV)

84,64%

Total fields extracted
(% over total elaborated OJV)

100,00%

Method

Release date
Structured fields collected
(% over total elaborated OJV)

100,00%

Total fields extracted
(% over total elaborated OJV)

100,00%

Salary
Structured fields collected
(% over total elaborated OJV)

20,54%

Total fields extracted
(% over total elaborated OJV)

20,60%

Method

Skill
Structured fields collected
(% over total elaborated OJV)

31,29%

Total fields extracted
(% over total elaborated OJV)

86,84%

Method

Working hours
Structured fields collected
(% over total elaborated OJV)

29,01%

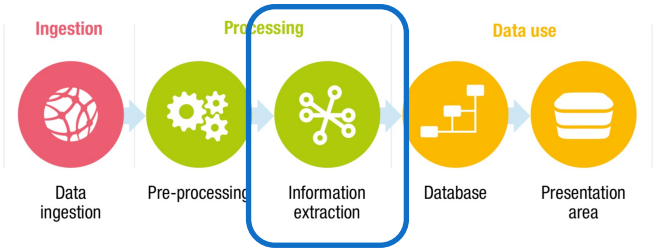
Total fields extracted
(% over total elaborated OJV)

63,91%

Method

Emsi | burning glass

Data Classification



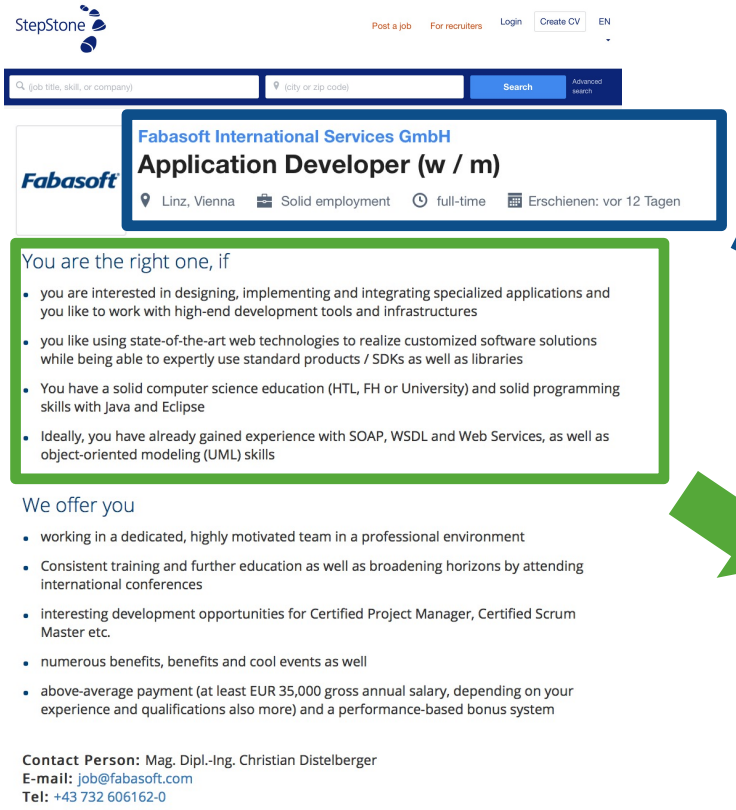
Extract and structure information from data,
with respect to the most proper taxonomy

Artificial Intelligence

Taxonomy selection

Information Linkage

Emsi | burning glass



Structured Data

Unstructured data

Information Extraction is an area of natural language processing that deals with finding factual information in free text.

This task uses machine learning techniques (ontology based learning, supervised learning and unsupervised learning) to match job ads with standard classifications.

Data cleaned and summarized

Staging Area

Occupation

Skills

Industry

…

Structured Data

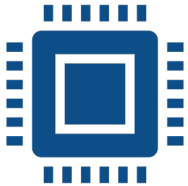
Staging Area

Machine Learning → Ontology based learning, supervised learning and unsupervised learning, NLP, etc.

Data classification
The recipe

Gold dataset

A training dataset is a dataset of examples used for learning



Classifier

The classifier is the concrete implementation of an algorithm that implements classification

Metric

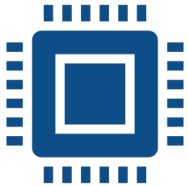
The metric that you choose to evaluate your machine learning model

Emsi | burning glass

Data classification
Occupation

Gold dataset

~150 / 200 k records by each language



Classifier

Ontology based Model + Naive Nayes Classifier

Metric

Weighted Precision

Emsi | burning glass

Workshop time @Databricks

https://github.com/mauropelucchi/sgi_workshop_2022

Emsi | burning glass

Databricks

- E’ una piattaforma di analytics facile e collaborativa basata su Apache Spark
- Tutti i moduli di Spark sono presenti in Databricks (SparkSQL, Streaming, ML, GraphX)
- “Main goal: to remove all the hardness and complexity to get and manage a Spark cluster”
- Permette di gestire installazioni e impostazioni con un clic
- Offre flussi di lavoro semplificati e area di lavoro interattiva per facilitare la collaborazione tra data scientists, sviluppatori e business analysts
- Integrazione con i maggiori cloud providers come Amazon AWS and Microsoft Azure

Notebook

- Simile ai notebook di Jupyter o Zeppelin.Linguaggi supportati: R, Python, Scala e SQL
  - Possono essere usati tutti in un singolo notebook.
- La Spark session è gia definita per ciascun notebook come variabile globale spark.
- Una volta creato un notebook deve essere collegato a un cluster attivo

Version Management

- Databricks è una piattaforma di analisi collaborativa dove gli utenti possono condividere workspaces, clusters and jobs attraverso una sola interfaccia.

- E’ possibile creare modelli condivisi nello stesso notebook real time, riutilizzare data assets, librerie presenti su uno stesso cluster, o riutilizzare/monitorare scheduled jobs.

- Databricks supporta l’integrazione con Github, Bitbucket Cloud & Azure DevOps Services

Data Powers AI and Machine Learning

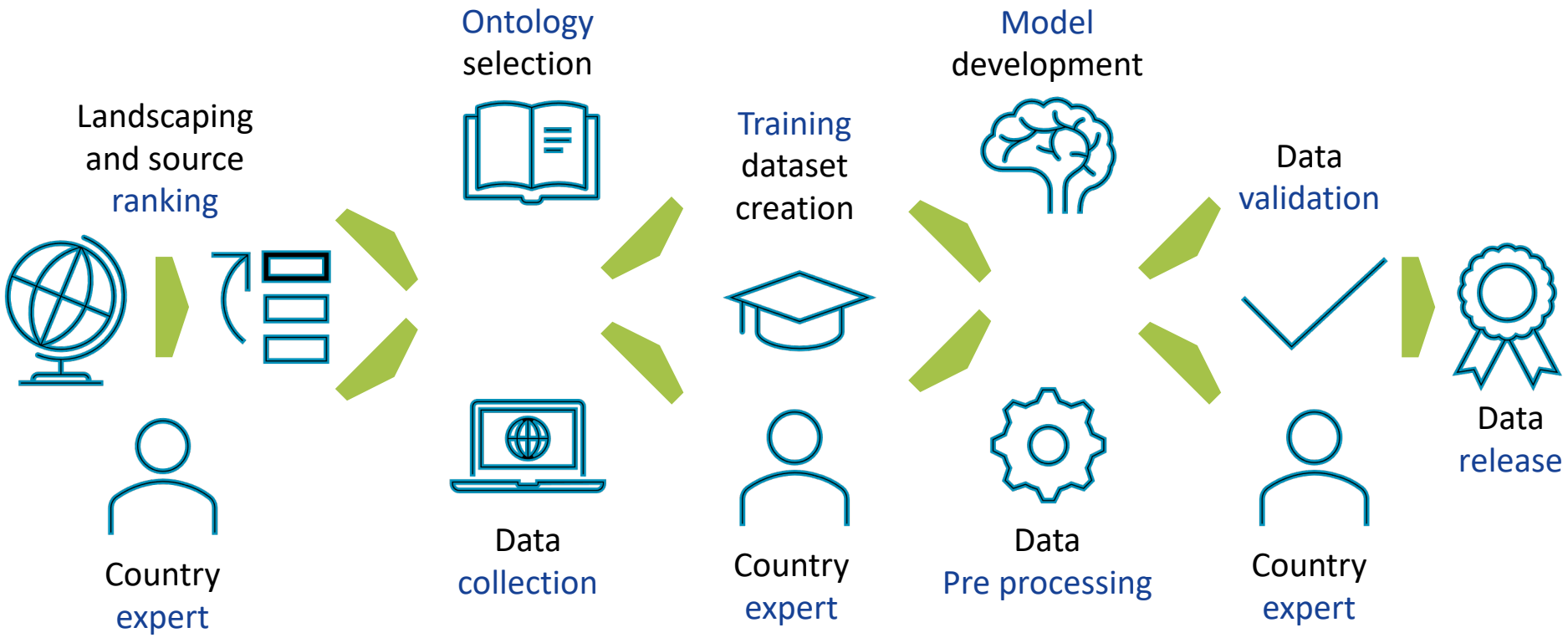
AI models can only learn from diverse, high quality data

More sophisticated models require even more data (high quality data!)

e.g. deep learning requires millions of documents!

Emsi | burning glass

The project pipeline



Emsi | burning glass

Improve ontology
Topic modelling and unsupervised learning

SAP Consultant

Posted 4 days ago by Ascendant Recruitment

Milton Keynes, Buckinghamshire
£50,000 - £60,000 per annum
Contract, full-time
Be one of the first ten applicants

SAP Consultant

12 month contract

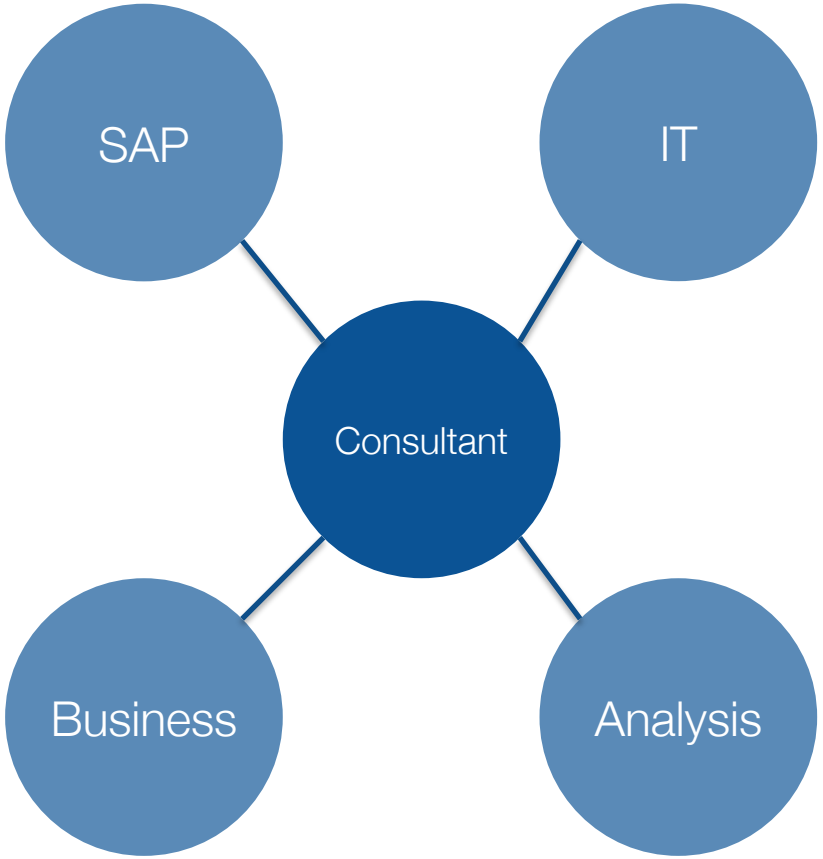
£50,000 - £60,000

Milton Keynes

We are delighted to be working with a prestigious client based in Milton Keynes for the recruitment of a qualified SAP consultant.

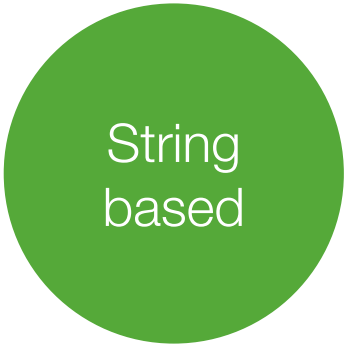
This is an exciting and challenging opportunity for an experienced SAP Consultant to join the e-business division of this large organisation.

The role will involve analysing business requirements as well as taking the lead in the design and delivery of solutions for various IT projects and business change programmes.



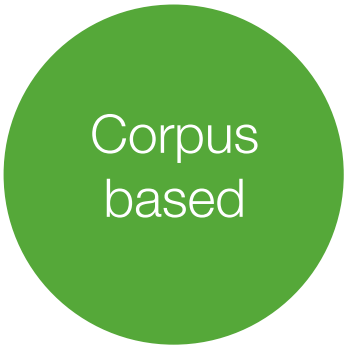
Emsi | burning glass

Text Similarity Approaches



String similarity measures operate on string sequences and character composition.

Jaro-Winkler, Jaccard, Cosine similarity



Corpus-Based similarity is a semantic similarity measure that determines the similarity between words according to information gained from large corpora.

Latent Semantic Analysis, Explicit Semantic Analysis, DIStributionally similar words using CO-occurrences

Knowledge-Based Similarity is based on identifying the degree of similarity between words using information derived from semantic networks

Emsi | burning glass

StepStone

Post a job | For recruiters | Login | Create CV | EN

(job title, skill, or company) | (city or zip code) | Search | Advanced search

Fabasoft International Services GmbH

Application Developer (w / m)

Linz, Vienna | Solid employment | full-time | Erschienen: vor 12 Tagen

You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

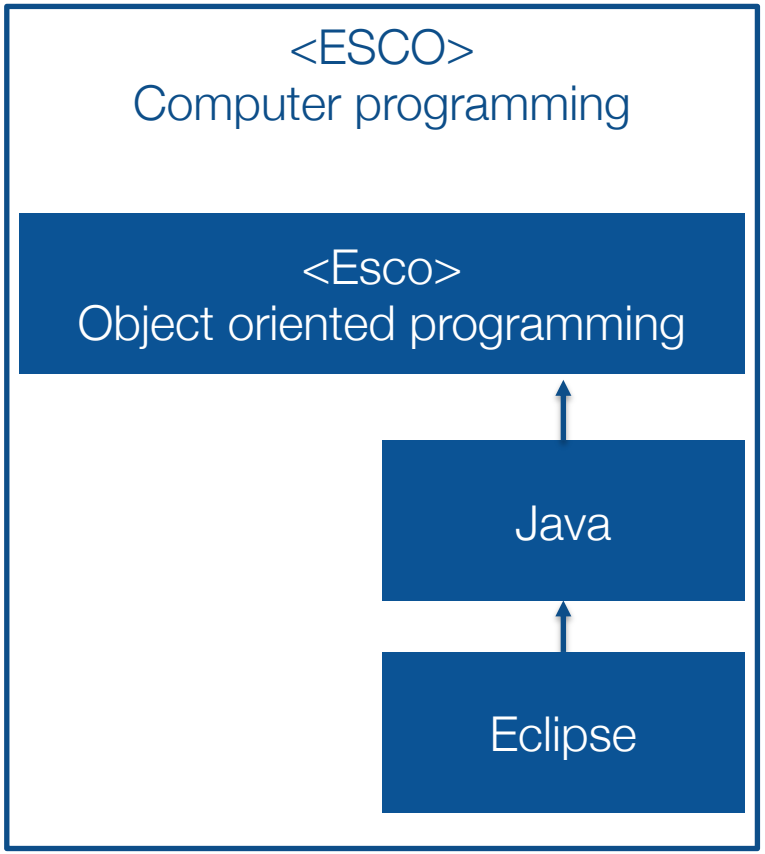
We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0

Consistent and linkage

<java> <eclipse>



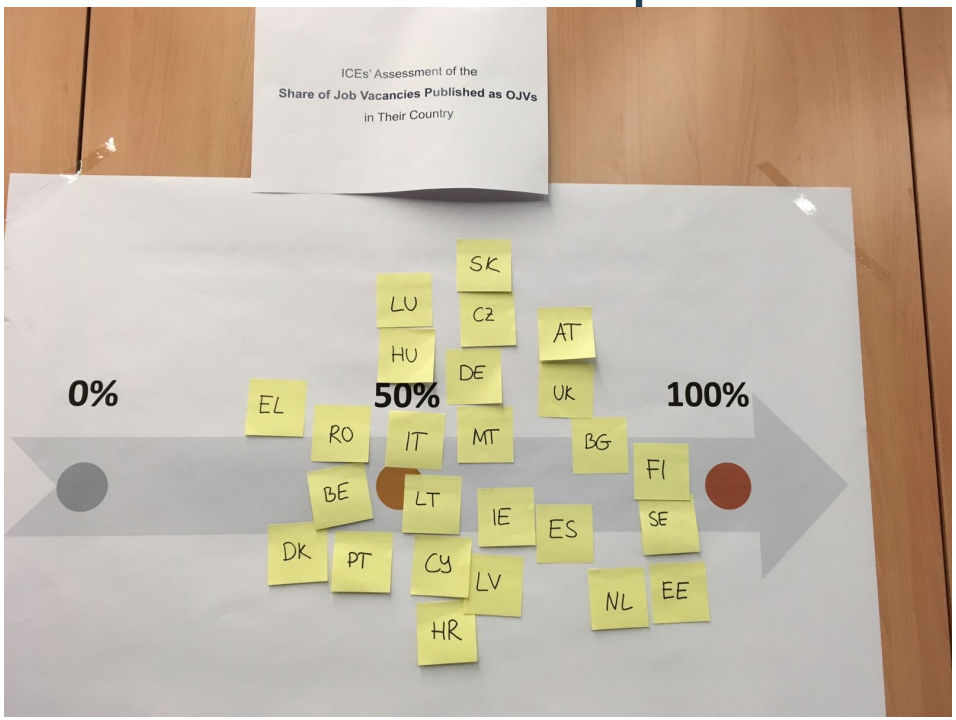
VALIDATION EXERCISE

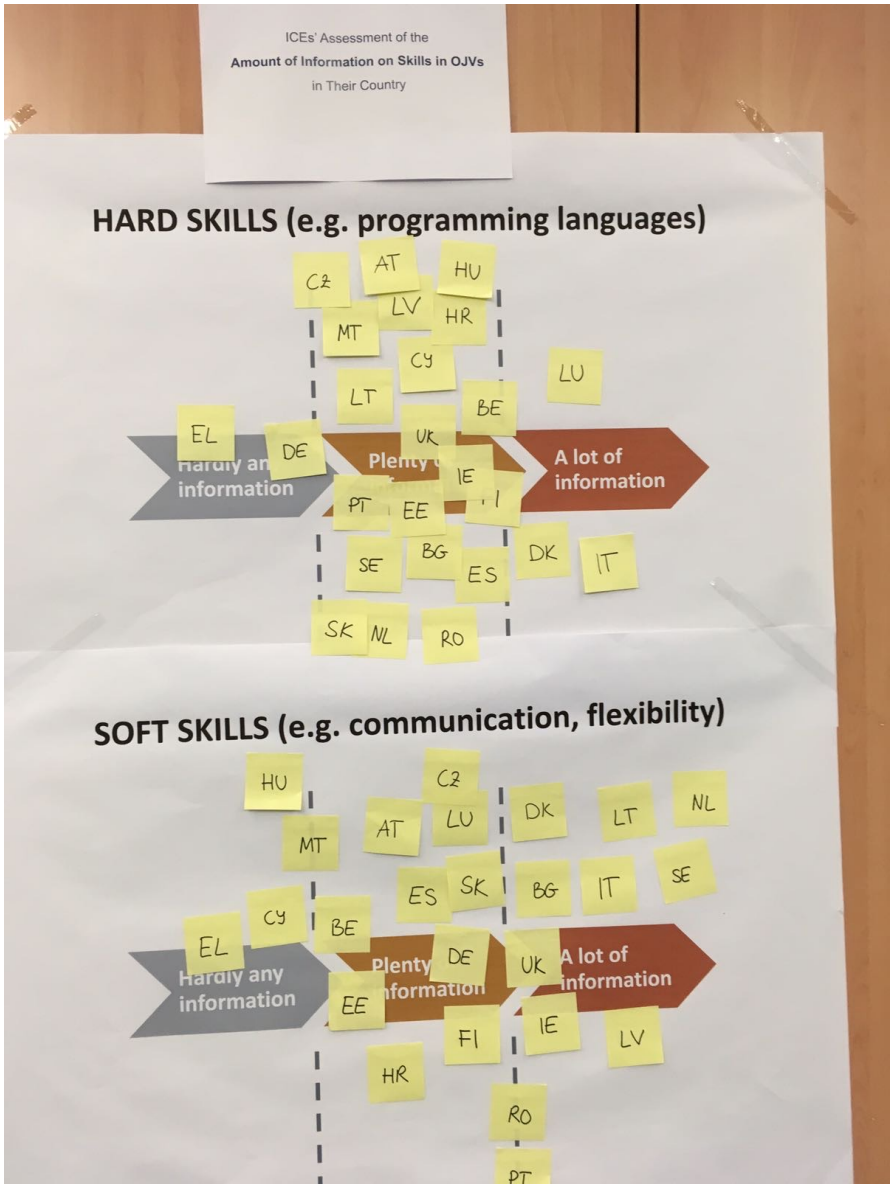
Emsi | burning glass

What do you think about "validation"?

Emsi | burning glass

Role of experts during Information Extraction and Classification process





Emsi | burning glass

System Validation

Validation phase in mainly composed by 2 processes:

- Comparison with official statistics
  - We don't expect to have totally corresponding distributions:
    - Some sectors could be underestimated (e.g. public administration or heathcare)
    - Some professions could not find a proper correspondence in official statistics (e.g. emerging occupation)
- International Country Experts
  - We engage country experts to help us validating data, due to:
    - Language knowledge
    - Specific Labour Market knowledge

Emsi | burning glass

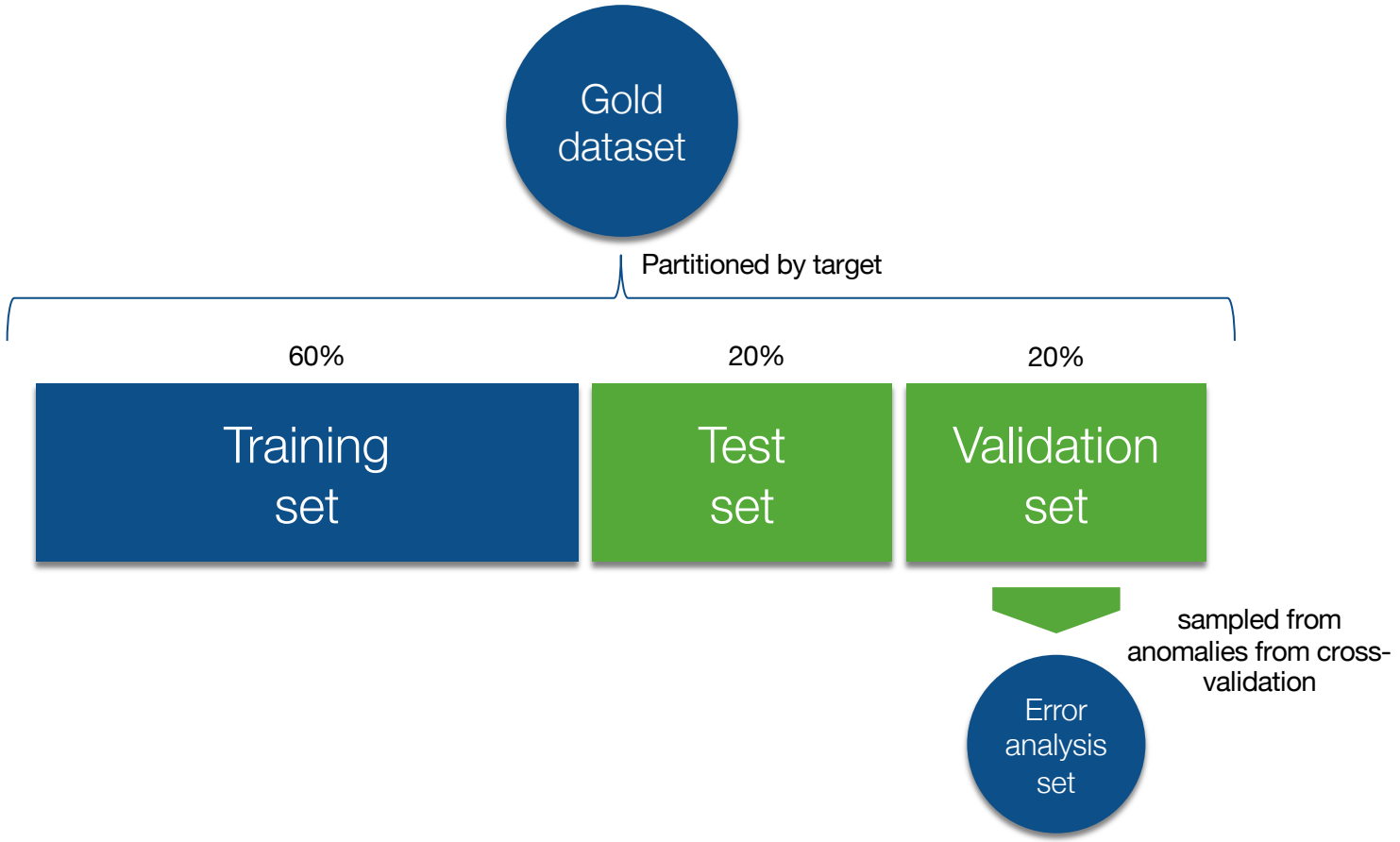
Data classification Challenges: Continous improvement

To guarantee a satisfying quality level of classification, we need to keep our models updated

It’s a Model Life-Cycle challenge

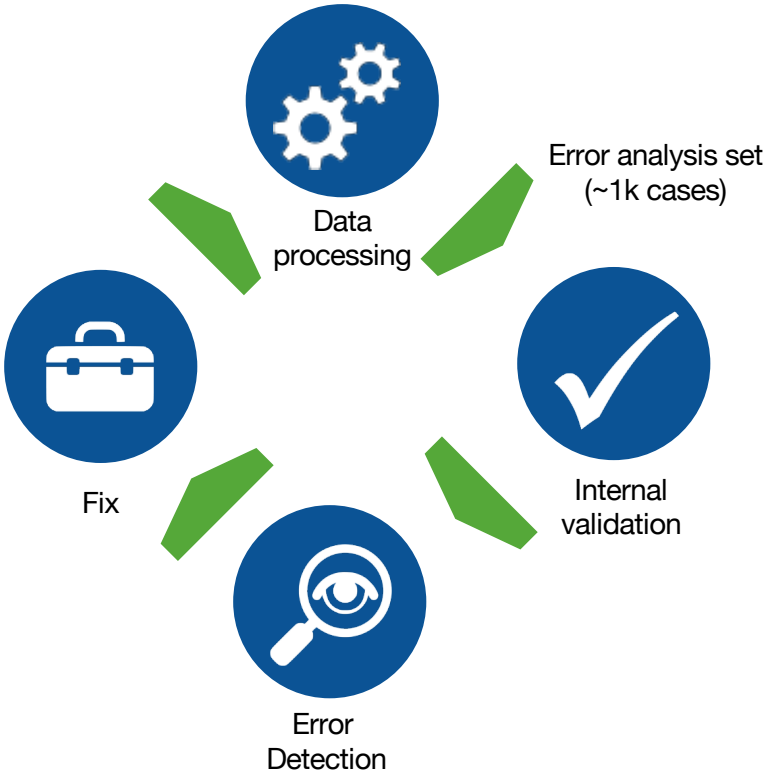
Let’s have a quick deeper look on in…

Model Life-Cycle
Error analysis set



Emsi | burning glass

Model Life-Cycle
Machine Learning diagnostic



Emsi | burning glass

Model Life-Cycle Recap

- Get more training examples
- Try smaller sets of features
- Add new features
- Try adding more complex features
- Check hyper-parameter of algorithm

More training examples:
Mark with OK/NO the records

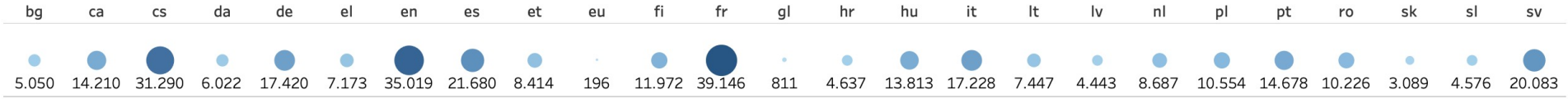
Find new labels and new features: fix an association between record and taxonomy

Emsi | burning glass

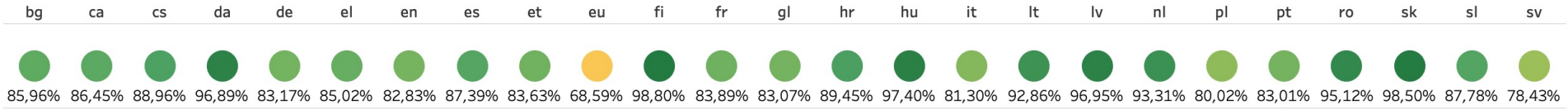
Precision of occupation (overall)

Validation Set (overall)

Validation Set by language



Precision of occupation by language



Precision of occupation (lv1)

Clerical support workers	85,77%
Craft and related trades ..	86,10%
Elementary occupations	86,19%
Managers	86,32%
Plant and machine operat..	86,29%
Professionals	86,61%
Service and sales workers	89,38%
Skilled agricultural, fores..	88,79%
Technicians and associate..	85,54%

Precision of occupation (lv2)

Administrative and comm..	85,06%
Agricultural, forestry and ..	80,82%
Assemblers	84,87%
Building and related trad..	92,30%
Business and administrati..	85,66%
Business and administrati..	80,06%
Chief executives, senior o..	91,36%
Cleaners and helpers	85,11%
Customer services clerks	82,21%
Drivers and mobile plant ..	86,49%
Electrical and electronic t..	74,60%
Food preparation assista..	89,08%
Food processing, wood w..	82,61%
General and keyboard cler..	97,20%
Handicraft and printing w..	89,65%

Precision of occupation (lv3)

Administration professio..	86,21%
Administrative and specia..	84,92%
Agricultural, forestry and ..	80,82%
Animal producers	83,13%
Architects, planners, surv..	87,56%
Artistic, cultural and culin..	91,74%
Assemblers	84,87%
Authors, journalists and li..	90,72%
Blacksmiths, toolmakers ..	86,70%
Building and housekeepin..	90,33%
Building finishers and rel..	95,47%
Building frame and relate..	90,00%
Business services agents	89,57%
Business services and ad..	79,10%
Car, van and motorcycle d..	90,40%

Precision of occupation (lv4)

Accountants	83,60%
Accounting and bookkeepi..	58,14%
Accounting associate prof..	85,65%
Actors	93,41%
Administrative and execu..	84,32%
Advertising and marketin..	65,30%
Advertising and public rel..	71,63%
Aged care services manag..	78,81%
Agricultural and forestry ..	94,55%
Agricultural and industria..	76,49%
Agricultural technicians	81,32%
Air conditioning and refri..	85,95%
Air traffic controllers	84,43%
Air traffic safety electroni..	95,52%
Aircraft engine mechanics..	79,61%

Emsi | burning glass

DATA ANALYSIS PATH

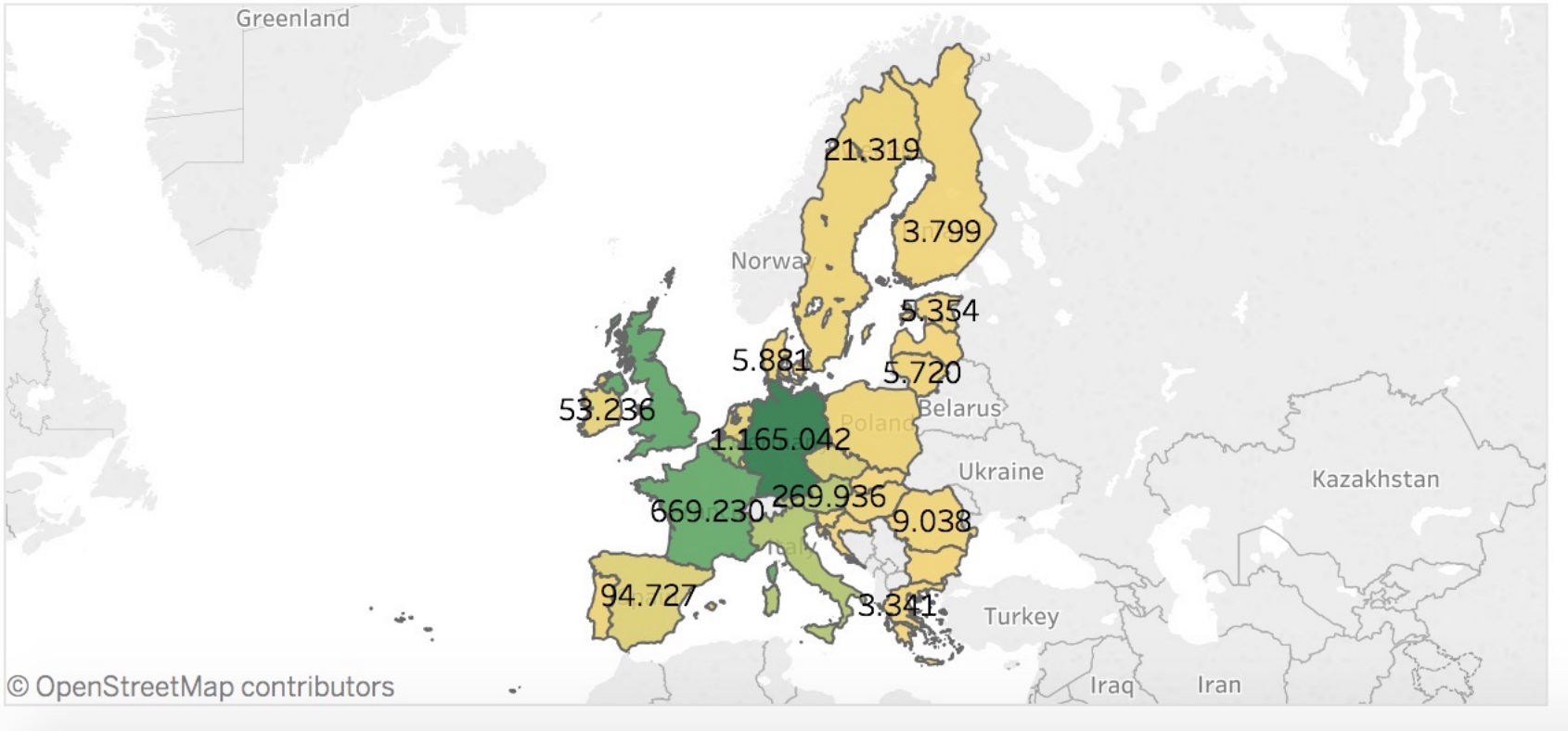
Emsi | burning glass

Number of vacancies

Number of sites

3.947.656

Vacancies by country



EMSIBG Team Whats-app group… at day 3 from the starting of the system

Challenges of Data Presentation – key points

Which Users and navigation patterns?

How data and users interact with the context?

How data must be integrated?

How to manage data governance?

Emsi | burning glass

Identify users and navigation patterns

Data
Scientists
and Analysts

Decision
Makers and
Business Users

Emsi | burning glass

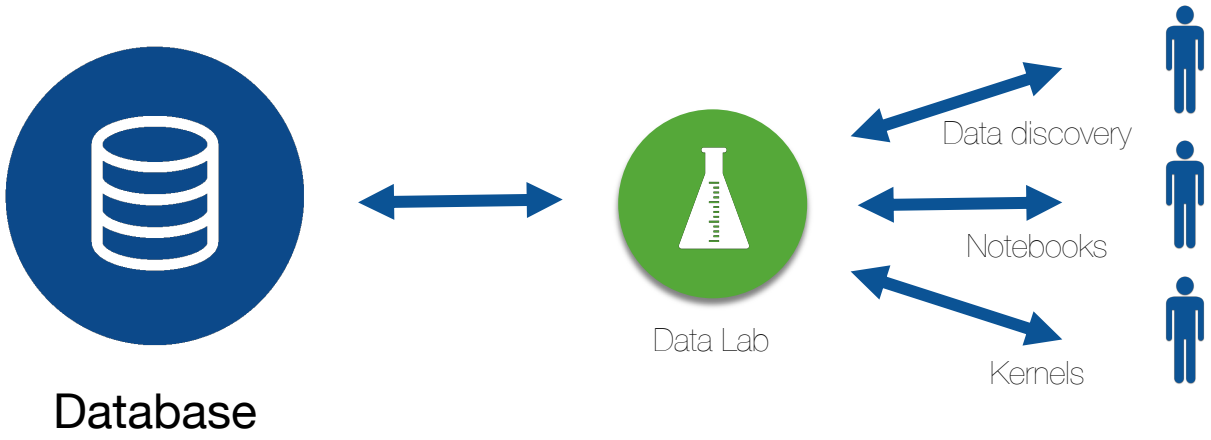
Data Scientists and Analysts

Data Discovery Capabilities

Publish, Share and Collaborate

Machine-Learning Integration

Embedded Advanced Analytics



Emsi | burning glass

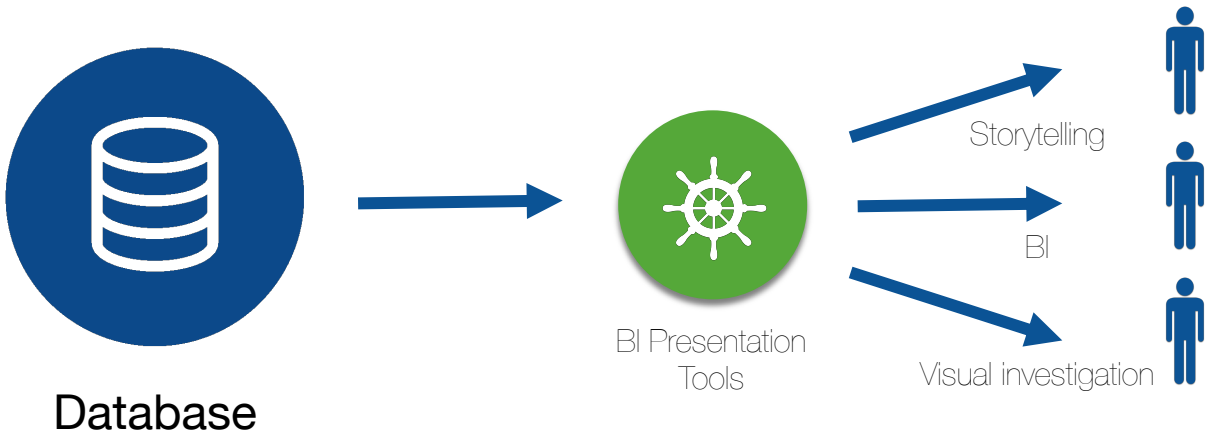
Decision Makers and Business Users

Self-service analytics and BI

Visual-Based Investigation

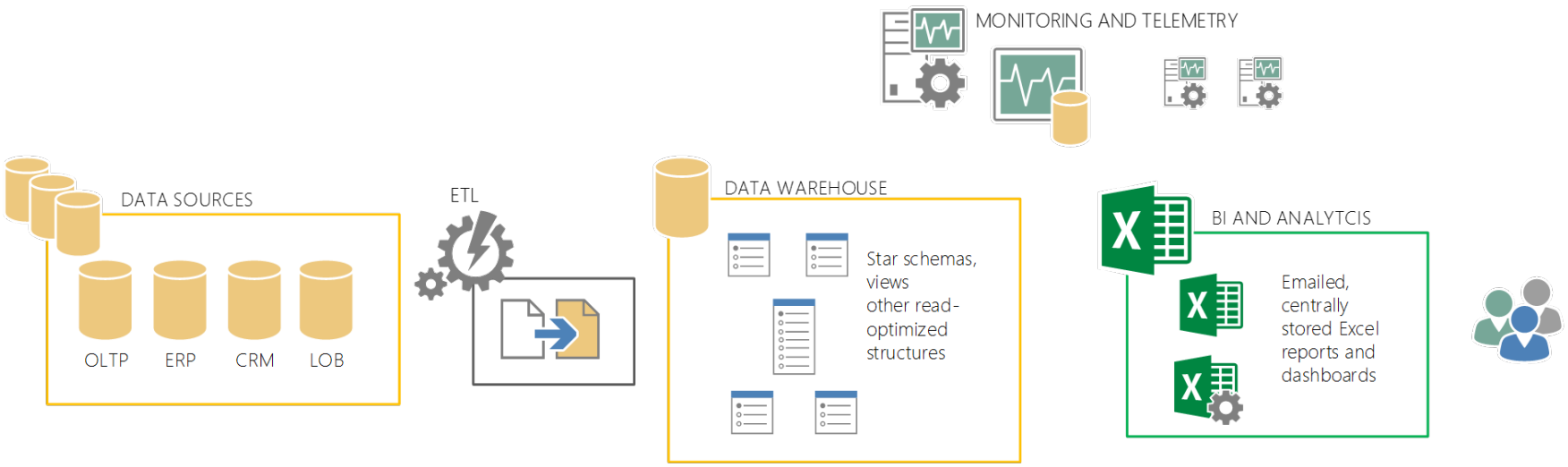
Governed Data Discovery

Data storytelling



Emsi | burning glass

Traditional approach



Well manicured, often relational sources

Known and expected data volume and formats

Little to no change

Complex, rigid transformations

Required extensive monitoring

Transformed historical into read structures

Flat, canned or multi-dimensional access to historical data

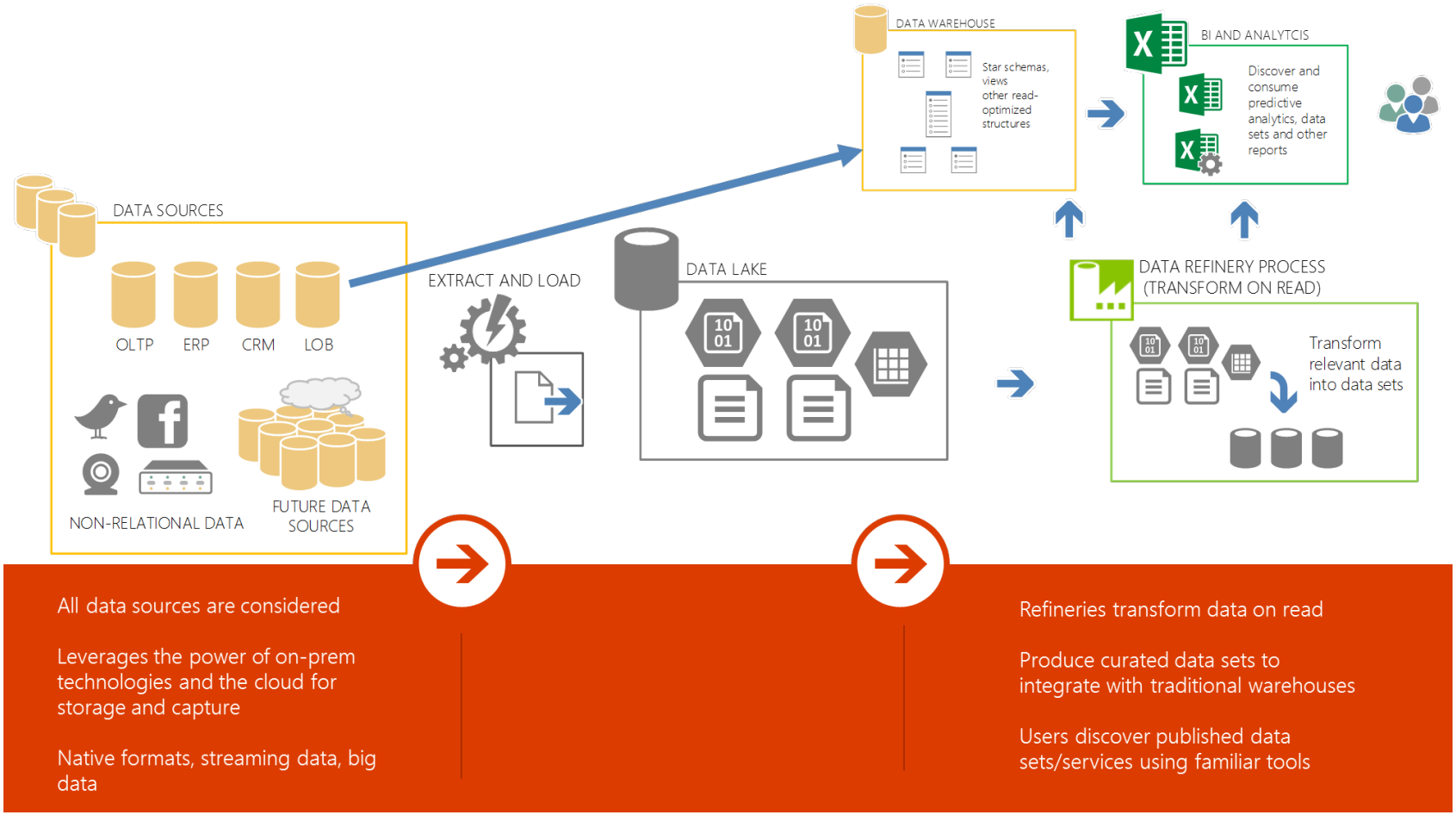
Many reports, multiple versions of the truth

24 to 48h delay

Credits: James Serra @ Microsoft

https://www.slideshare.net/jamserra/big-data-architectures-and-the-data-lake

Big Data approach



Credits: James Serra @ Microsoft

https://www.slideshare.net/jamserra/big-data-architectures-and-the-data-lake

Top Down + Bottom Up approach

What happened?

Descriptive Analytics

Why did it happen?

Diagnostic Analytics

What will happen?

Predictive Analytics

How can we make it happen?

Prescriptive Analytics

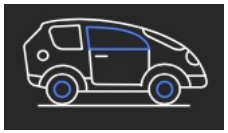
Credits: James Serra @ Microsoft

https://www.slideshare.net/jamserra/big-data-architectures-and-the-data-lake

Problem Statement

Automation
“Add data in minutes”

Scaling
“Support Exponential Growth”

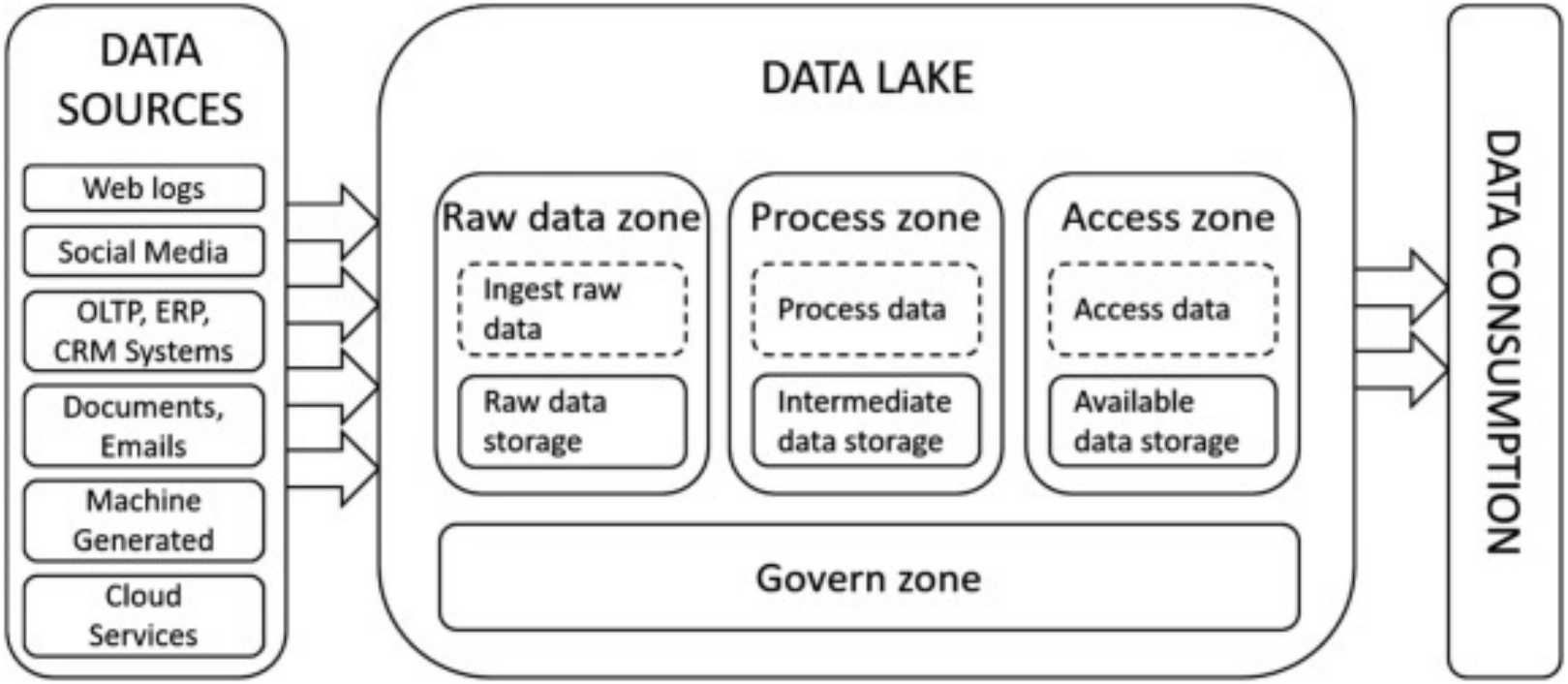


Any data, any sources

From Reporting to ML
“User configurable”

Accessibilty
“Tools and log-on”

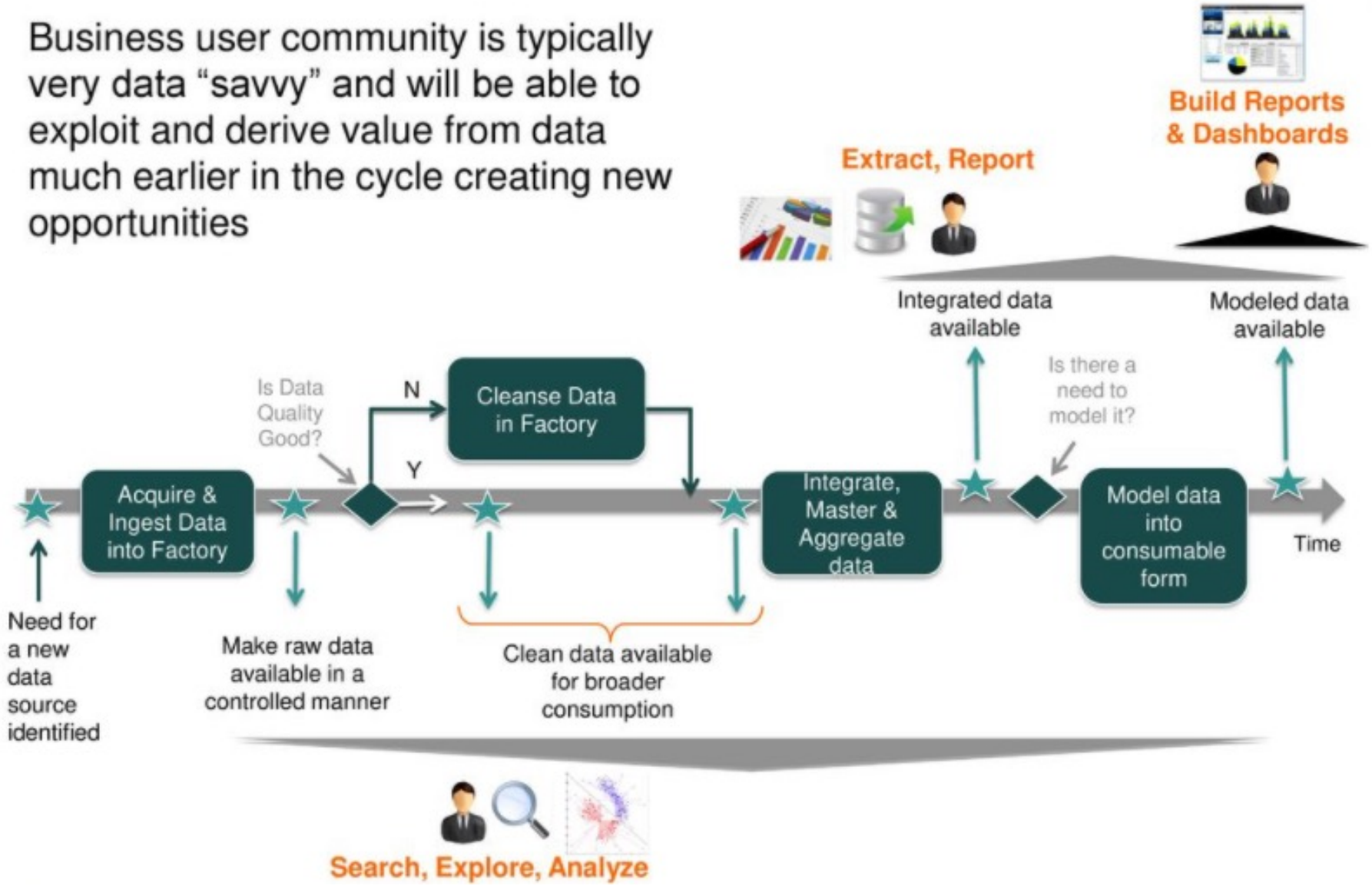
Emsi | burning glass



Emsi | burning glass

Value - Unlocking the Value of Data Earlier

Business user community is typically very data "savvy" and will be able to exploit and derive value from data much earlier in the cycle creating new opportunities



Modern Data Lake Architecture

- Schema-on-Read
- Descriptive Data Modeling
- New Data can start flowing any time and will appear retroactively
- Flexibility
- Scalability
- Rapid Data Ingestion
- Good for Exploration and Botton-Up Approach

Emsi | burning glass

DESIGN, IMPLEMENTATION AND EVOLUTION

Emsi | burning glass

Big Data Flow

Quality requirements

Micro-services design

Components by definition

Infrastructure challenges

Key design projects

- Micro-services
- Componentization
  - Component specialization
  - Small applications
  - Portability
  - Reuse
  - Maintenance
- Scale Out
  - Performance

Emsi | burning glass

Microservices

Language Detector	Spam Filter	No-Vacancy Filter	Stemmer	Deduplication component
TF-IDF Transformer	N-gram component	Text Cleaner	Merge Vacancy	Document2Vec
Tokenizer	StopWords Removers	Skills Classifier	Occupation Classifier	Education Requirements Classifier
Industry Classifier	Salary Extractor	Experience Extractor	Dates Extractor	WorkingHours Detector
Contract Detector	Locations Detector			

Emsi | burning glass

Organize around business services

Language Detector

Occupation Classifier

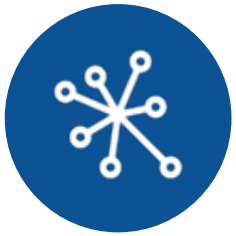
Salary Extractor

Skills Classifier

Emsi | burning glass

Evolution of the system

New Sources



New Machine Learning Models

New taxonomies

New dimensions

Increase Volume

New Insights

Deeper Knowledge

More Information

Emsi | burning glass