Pisa, 10 July 2019

Developing data services for audit

Magdalena Cordero

Director of Information, Workplace and Innovation

EUROPEAN COURT OF AUDITORS

In 2015 the ECA

Scenario 2

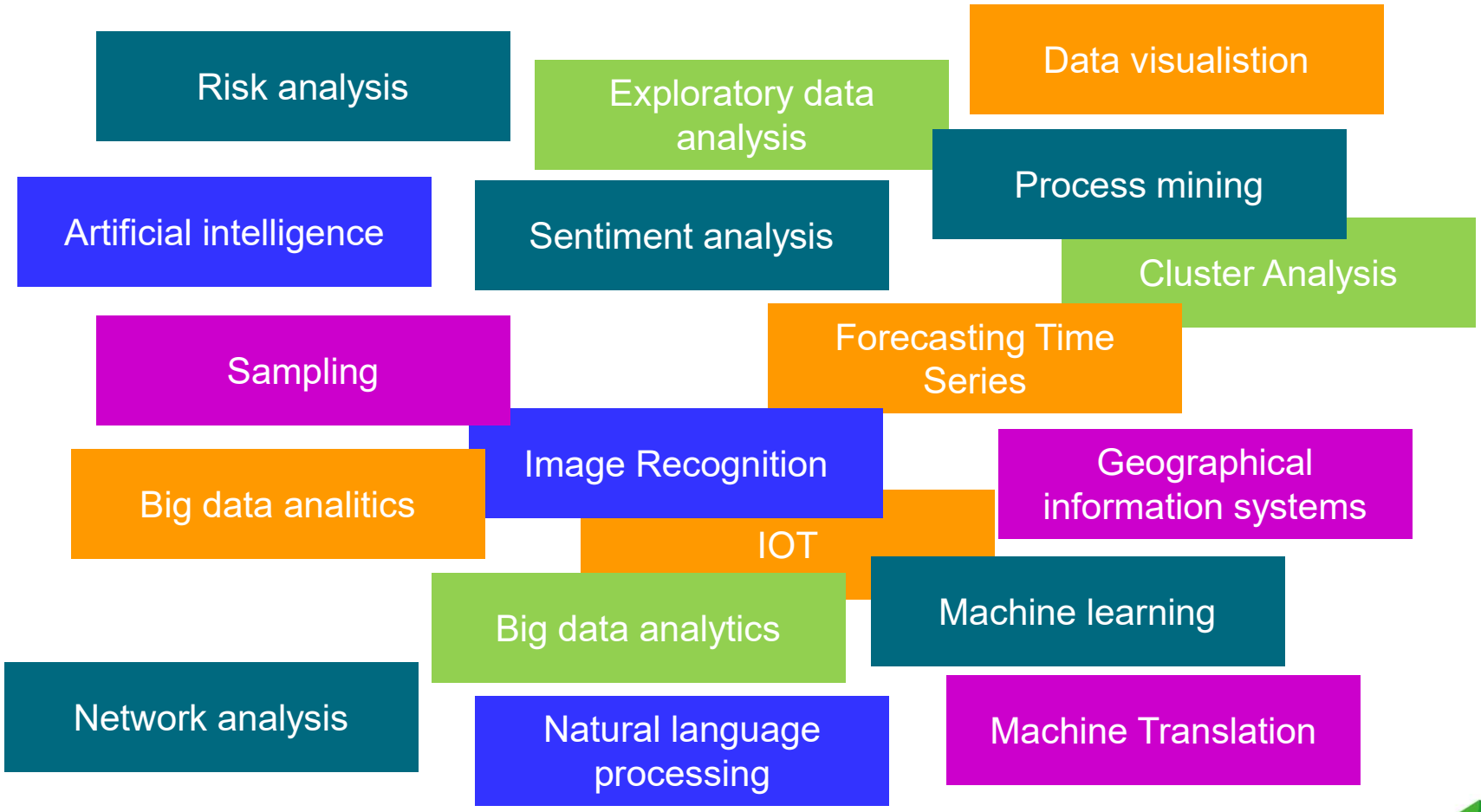
In **2015**, …

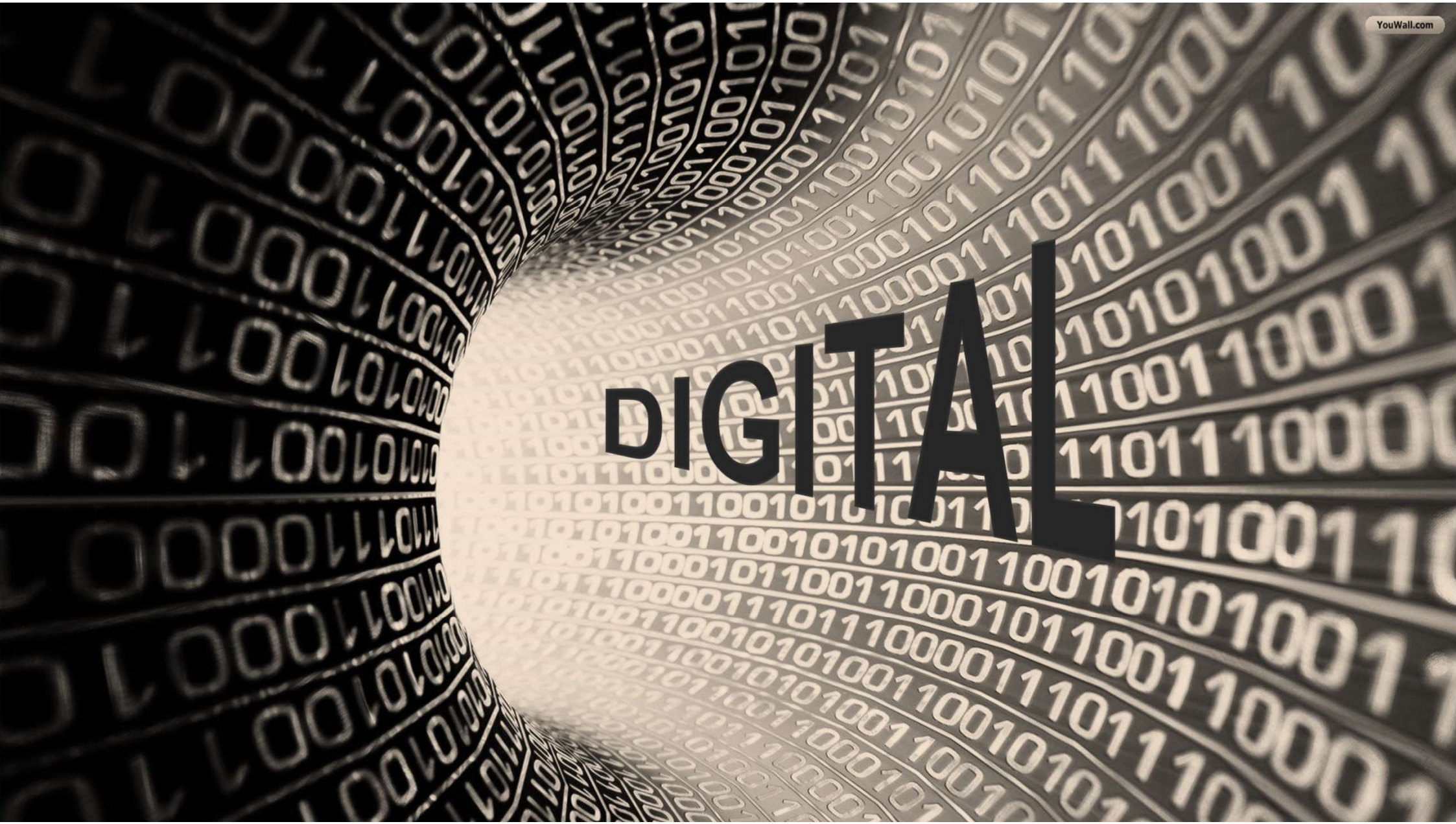
… the generation and availability of large amounts of data is transforming the world's economies ("big data"). At the same time, access to sophisticated data analysis methods is becoming cheaper.

Therefore, in **2040** …

… the ECA is using a high degree of automation in its audit procedures. Algorithms spot irregularities in digitalised reports and documentation. Artificial intelligence detects performance patterns in large data sets. Auditors will focus more on asking the right audit questions than on verification and analysis.

Techniques and methods



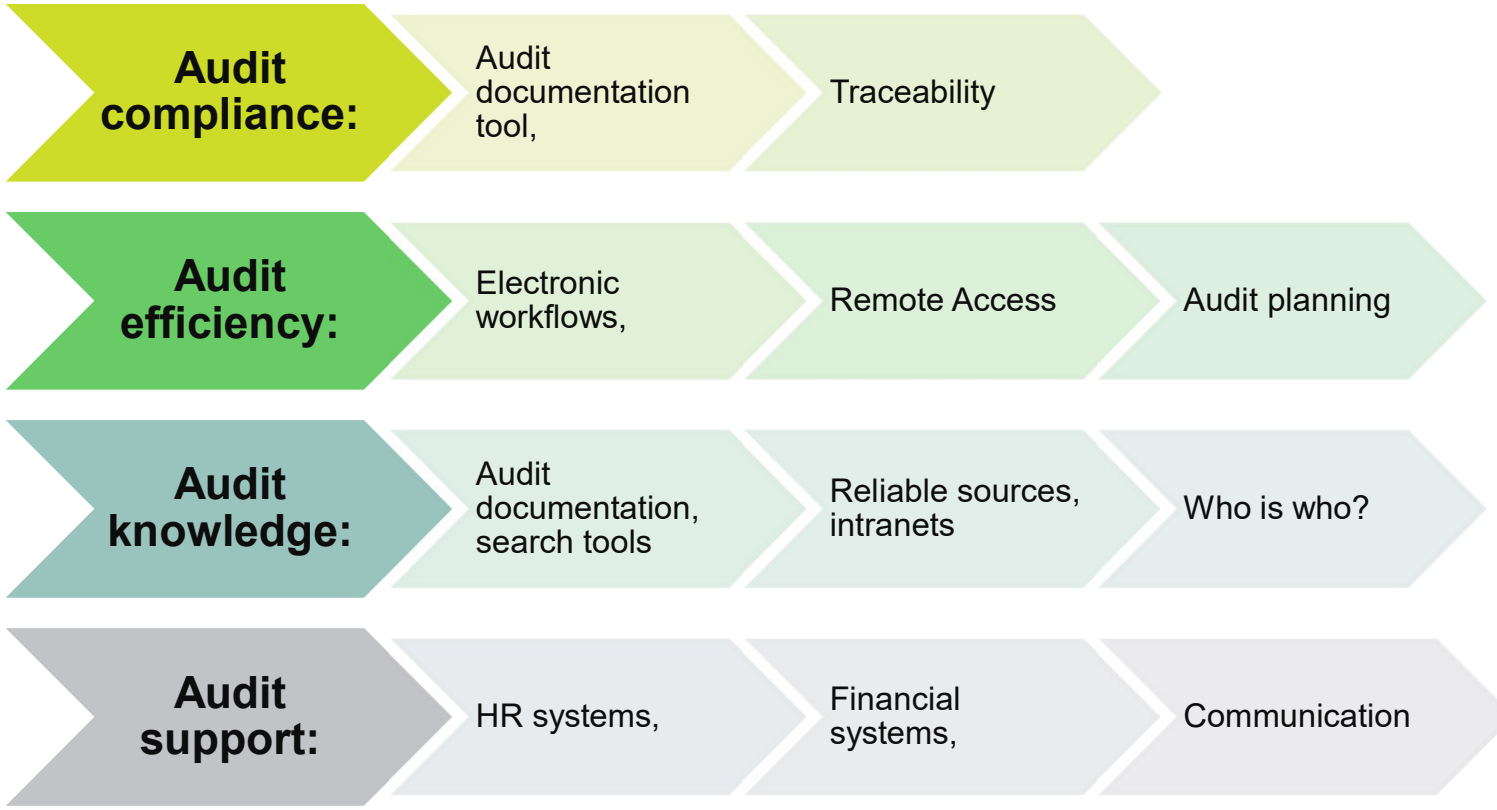

YouWall.com
DIGITAL

Digital transformation, digital audit

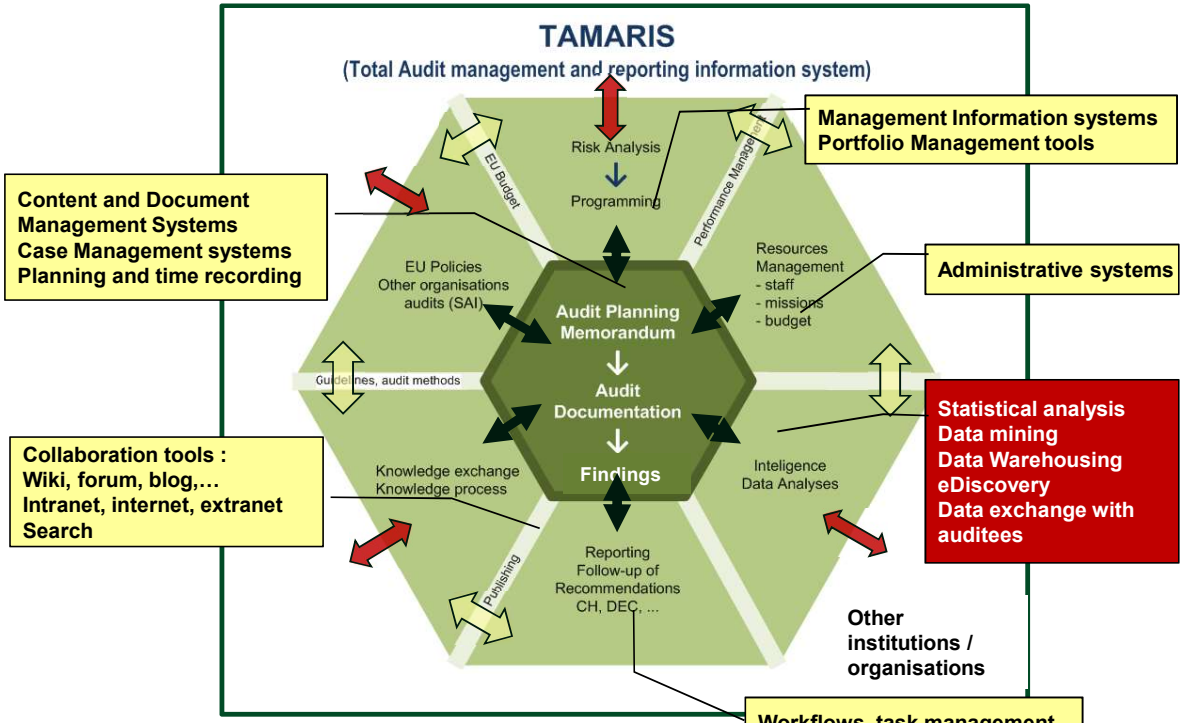
«The profession needs to achieve a "quantum leap" to redesign audit processes using today's technology, rather than using information technology to computerise legacy audit plans and procedures.»

AICPA. Reimagining Auditing in a Wired World
Vasarhelyi et all

Technology for audit: traditional role



IT supporting existing processes



Technology for audit: traditional role

IT audit

COBIT

- Control Objectives for Information and related Technology
- Since 1996. ISACA.
- Version 5

Next step

Audit transformation

From HOW to WHAT – changing the audit process

Why now?

«Measure what is measurable and make measurable what is not so.»

Attributed to Galileo Galilei

Sometimes you don’t have
the wrong idea,
just the wrong timing

Lady Norman

on her Krupps scooter in 1916

Audit transformation

Why?

- Evidence is digital
- Presence is digital
- Auditees are more and more digital

Technology maturity and availability

- Data exchange methods
- Data analytics
- Process mining
- Big data
- Natural language processing
- AI
- Machine learning

The evidence is DIGITAL

- Digital financial processes
- Digital administrative processes
- Electronic signature
- E-Government
- e-tendering, e-invoicing
- GPS
- Electronic charts
- **Digital twins/proxies**

The presence is DIGITAL

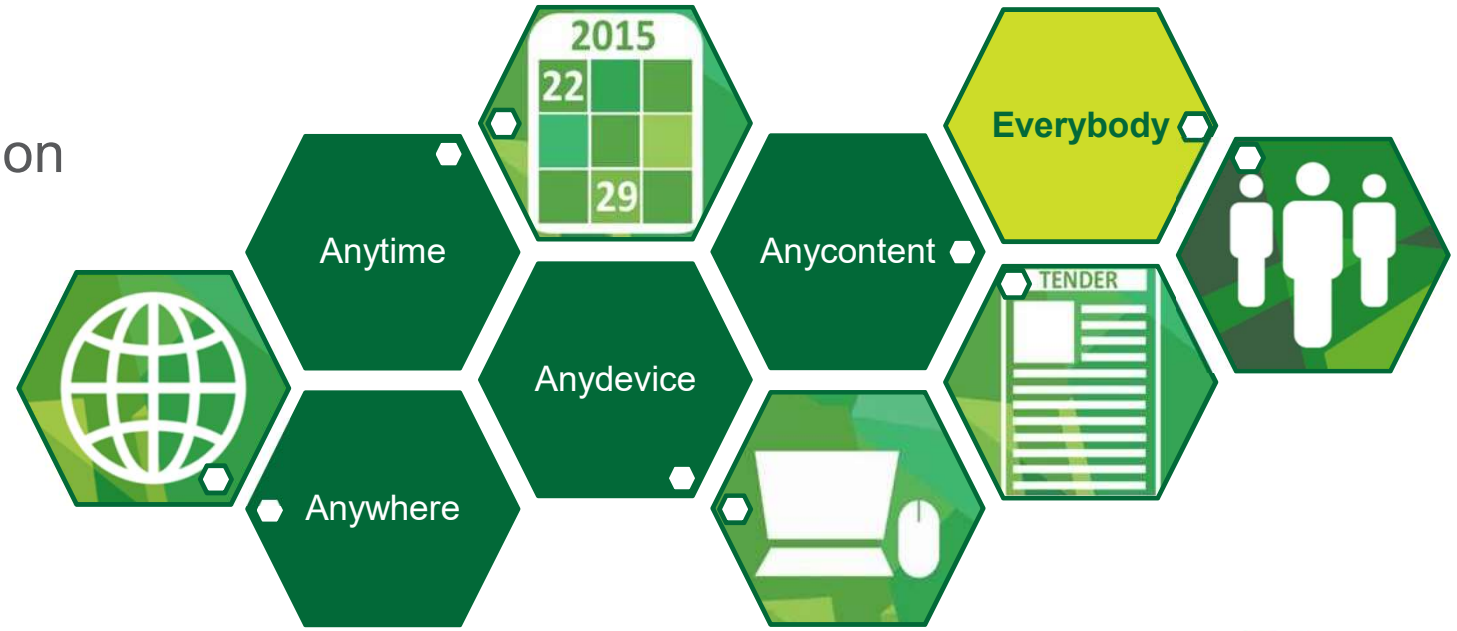
Work from **Anywhere**

At **Anytime**

Using **Anydevice**

To access **Anycontent**

Everybody: Collaboration



The control is digital

- Controls performed by the auditee documented electronically
- Controls perfromed by the controller of the auditee (internal audit) also electronically documented
- The full control system is described on databases and electronic document

Digitalisation allows control by design

eGovernment

Tallinn declaration on eGovernment

- Digital-by-default, inclusiveness and accessibility
- Once only
- Trustworthiness and Security
- Openness and transparency
- Interoperability by default
- Horizontal enabling policy steps

New digital services: new data and new risks

Risks: coordination, interoperability, common architecture

The key element

Data

Audit institutions don’t produce the data needed to perform audit.

Changes in audit

«The best way to predict the future is to invent it»

Alan Kay

the ABCD of digital audit:

A for **A**rtificial Intelligence

B for **B**lockchain

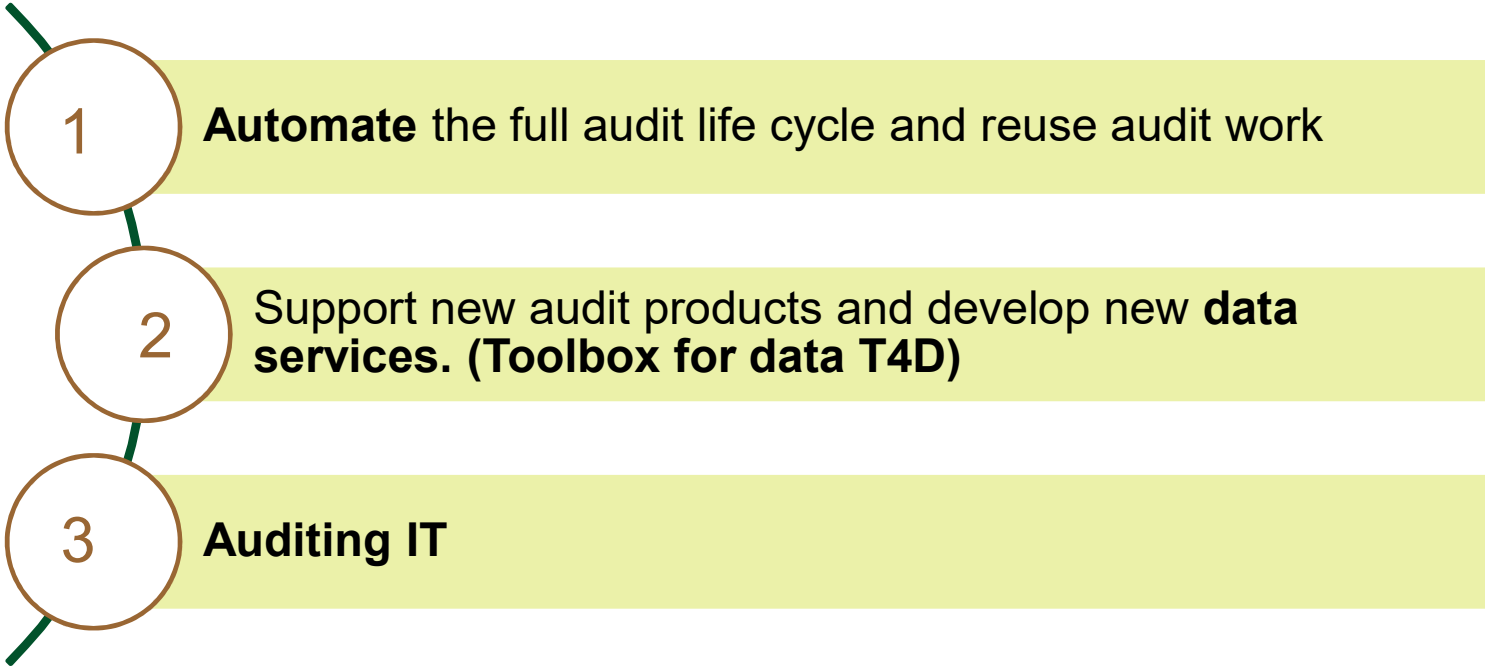
C for **C**yber-security

D for **D**ata Analytics

Robert Hodgkinson
ICAEW

Digital audit

3 dimensions



Audit Automation

- **Automation of the audit documentation process**
  - Single process for all information exchange
  - Blockchain for registering potential audit evidence: **control by design**
  - …
- **Automation of the activities of the auditor**
  - Data extraction and processing
  - RPA (robotic process automation)

  - Automation of the Financial audit of the Executive Agencies

Data Analysis

- Creation of a Toolbox for Data (T4D) and develop data services
- Identify potential use of data at the APM preparation phase
- Tools to transform the Statement of Assurance
- Support access to data sources (in particular European Commission data)

Data Analysis

Structured and unstructured information

Several Techniques:
Statistical, Data visualisation, Data mining, Process mining, Big data, Artificial Intelligence, Machine learning

For
- Compliance
- Policy evaluation
- Risk analysis (planning phase)
- Evidence gathering
- Having insights
- Fraud detection

Auditing IT

- Governance and procedures
- Information systems
- Controls implemented in the systems
- Algorithms (bias, values, ethics)

Roadmap – Towards digital audit

Short term:

RE-USE

analytics used to improve current processes

- Knowledge Node on Data science
- Identify case of application
- Training of auditors in analytics using existing tools

Mid term:

EXPERIMENT

identify digital impact on the audit process

- Create a Data Lab
- Recruit and train data scienctis
- Investigate improvements to Financial audit
- Investigate improvements to Performance audit

Long term:

INDUSTRIALISE

Propose new ways of working

- Redefinition of the processes
- Elaboration of new products
- Creation of an analytics tool set to be used autonomously by auditors
- Creation of cross-functional teams

The importance of sharing knowledge ECALab and ECALabers

The ECALab

Created in 2017 as a step towards establishing data services tailored for audit, the ECALab is:

with different job profiles and having personal interest in big data, data analytics, text mining, visualisation, machine learning, etc.

equipped with "up-cycled" IT equipment such as servers and laptops where tools can be tested in real life conditions

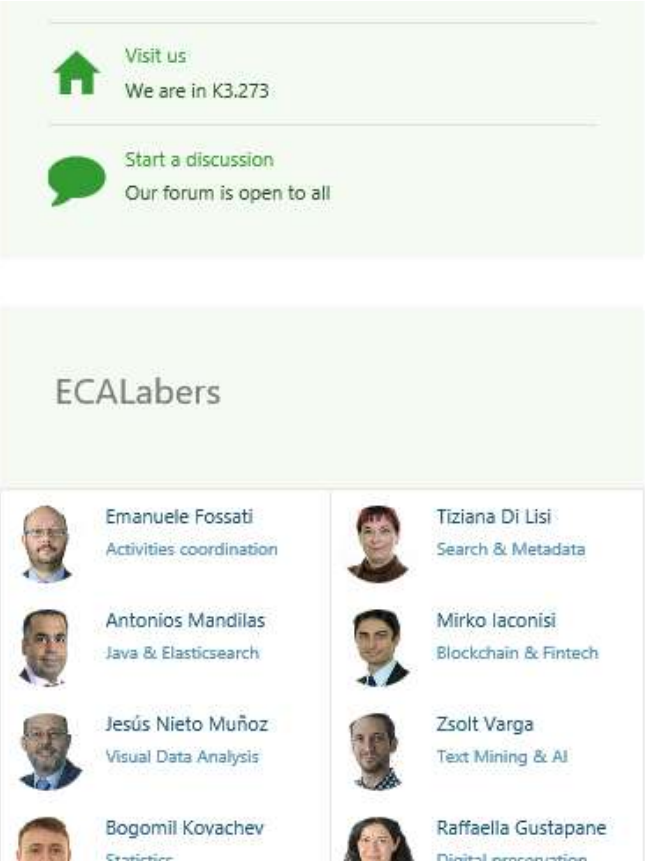
to share knowledge and ideas on technologies for audit

where they can come to ask for advice, share ideas and search for solutions to problems or questions they have

with books and people who are available to explain

ECALab: A physical space

ECALab home page

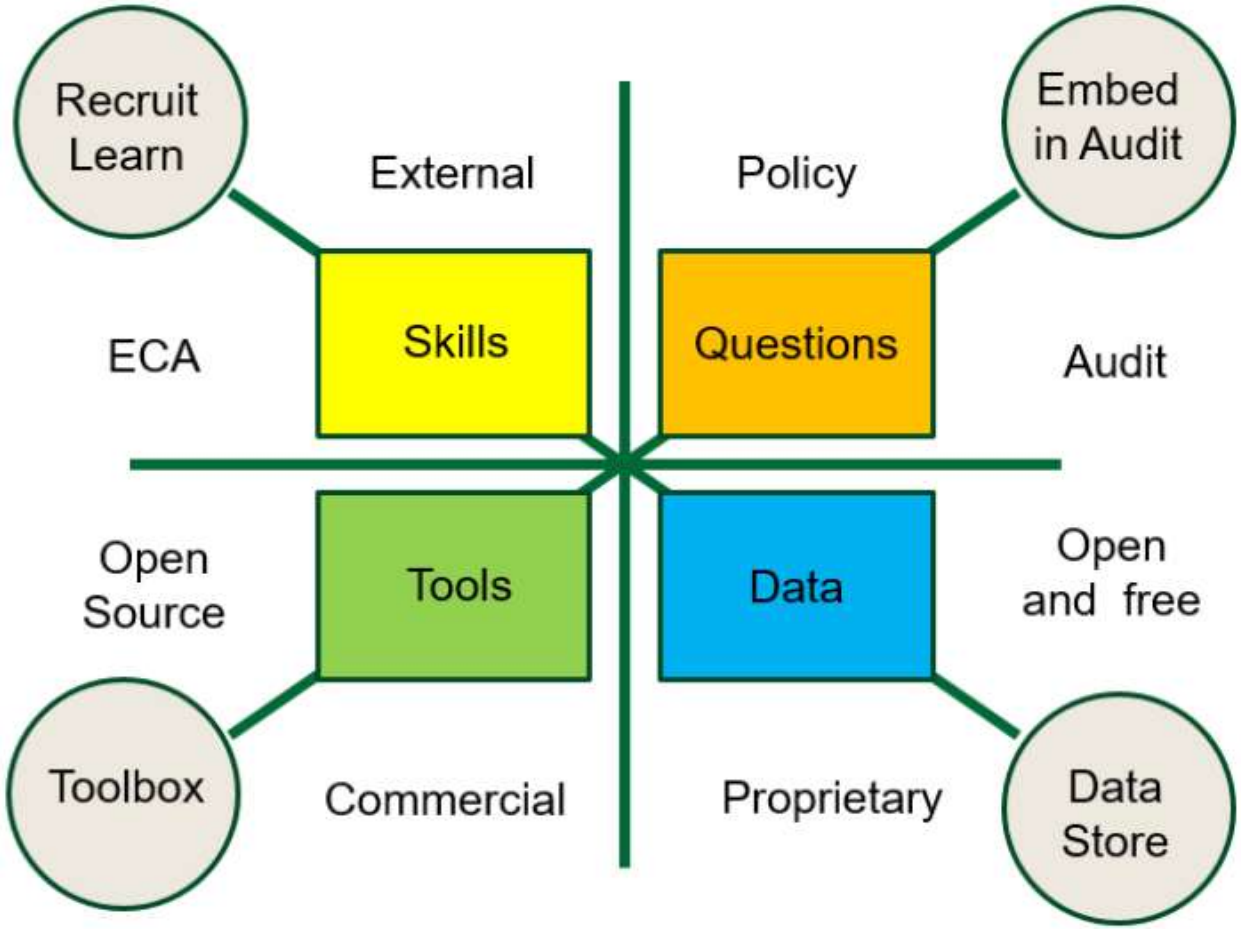


A framework to develop data services

If you know what you are doing, you are not doing research.

M. Vasarhelyi

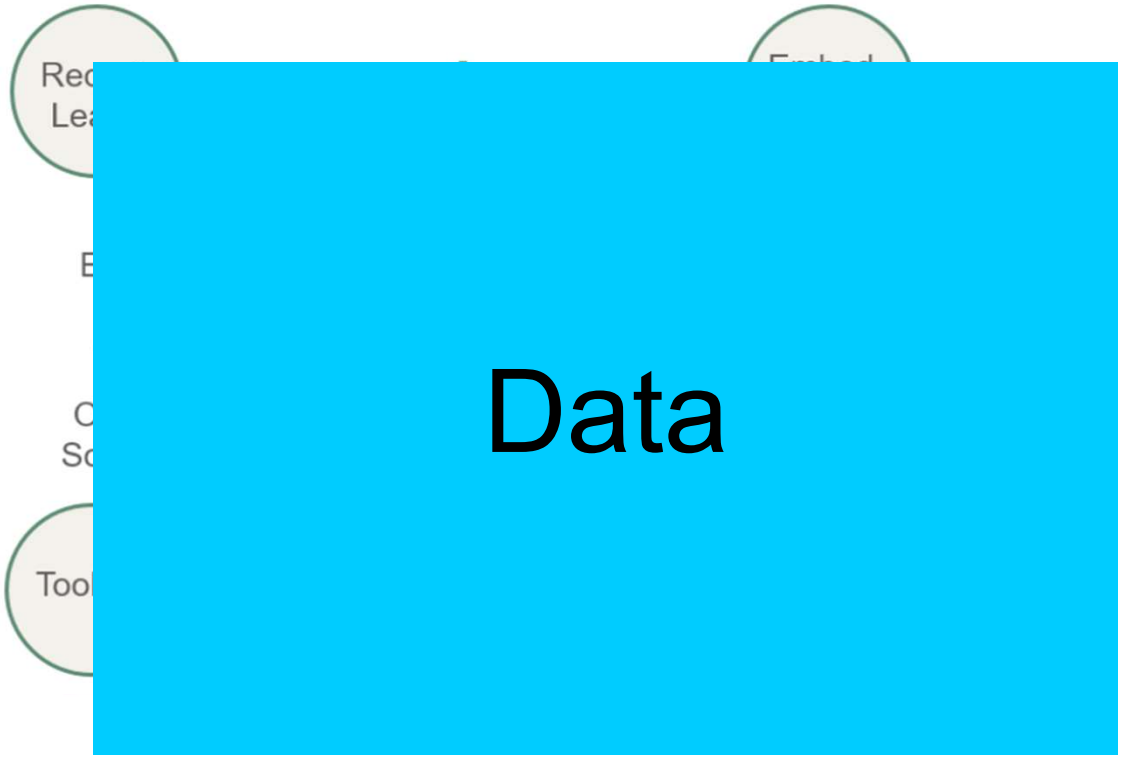
A framework to develop data services



EUROPEAN COURT OF AUDITORS

Why a framework for audit?

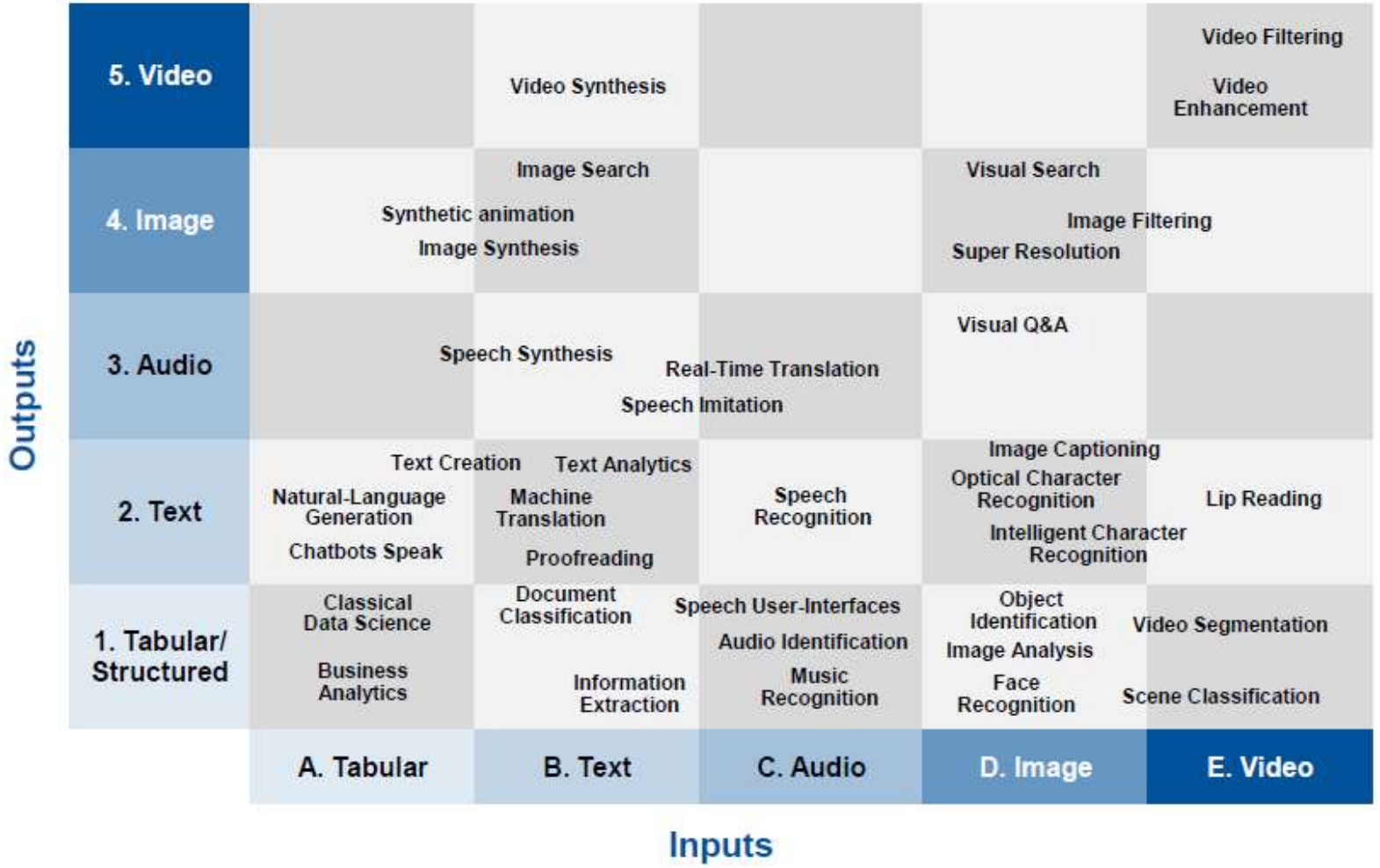
- Audit organisation aren’t data producers
- The planned time for audit is limited
- Components should be reusable and reused
- Each audit is like a new project


Data

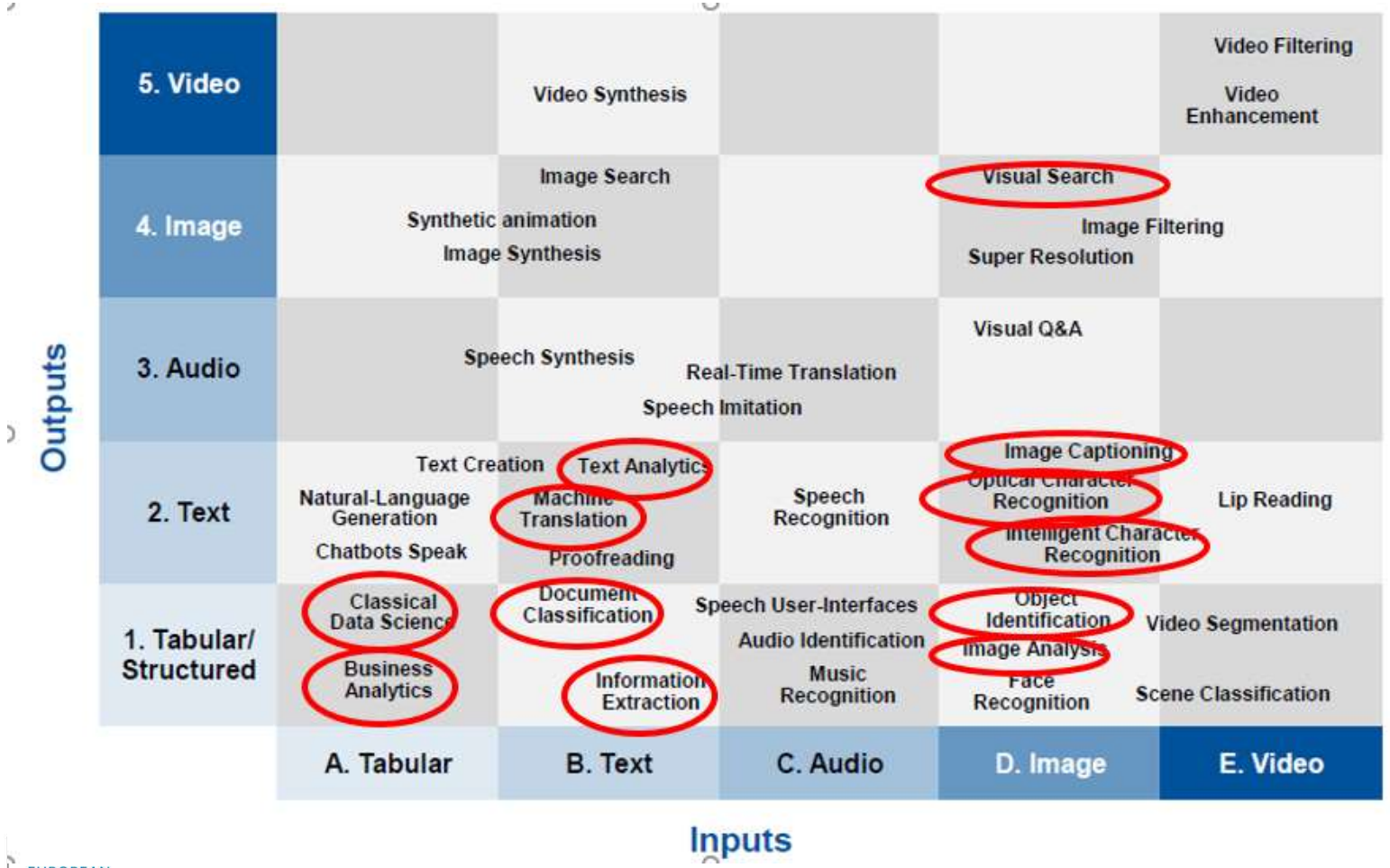
Data

- Identify **reliable data sources**
- Create a data sources catalogue : quality, access mechanism
- **Structured and unstructured data** : 90% of the data available on the internet are unstructured!
- Public data / Open data
- Proprietary data / Data brokers
- Data must be combined to generate knowledge
- Big data : volume, variety, veracity, velocity, value,… ?
- Media and social networks data
- Create a data store

Data formats, tools and transformations



Technologies to transform the data

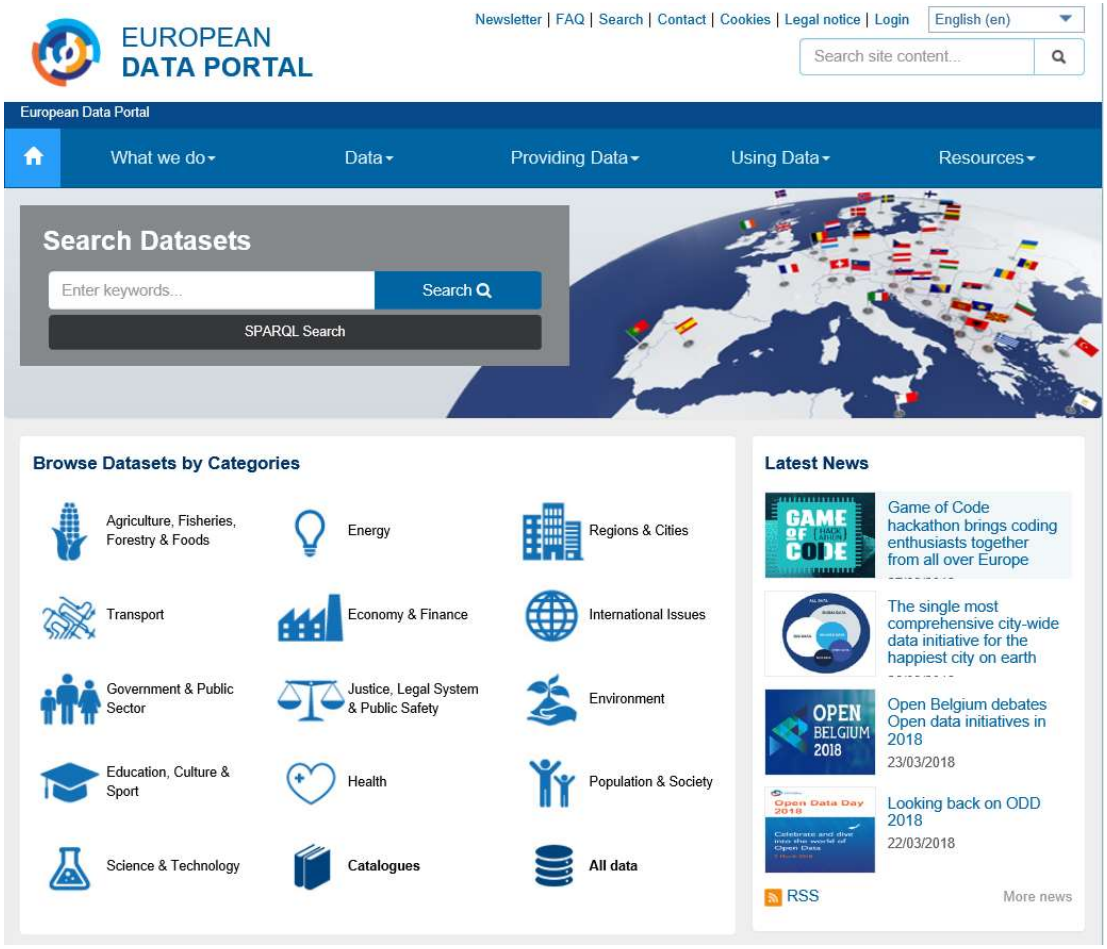


Open data

- **Public Sector Information** (PSI) is information collected, produced or paid for by the public bodies
- **Open government data** is PSI made freely available for re-use for any purpose.
- EDP harvests the metadata of open government data available on **public data portals** across European countries.

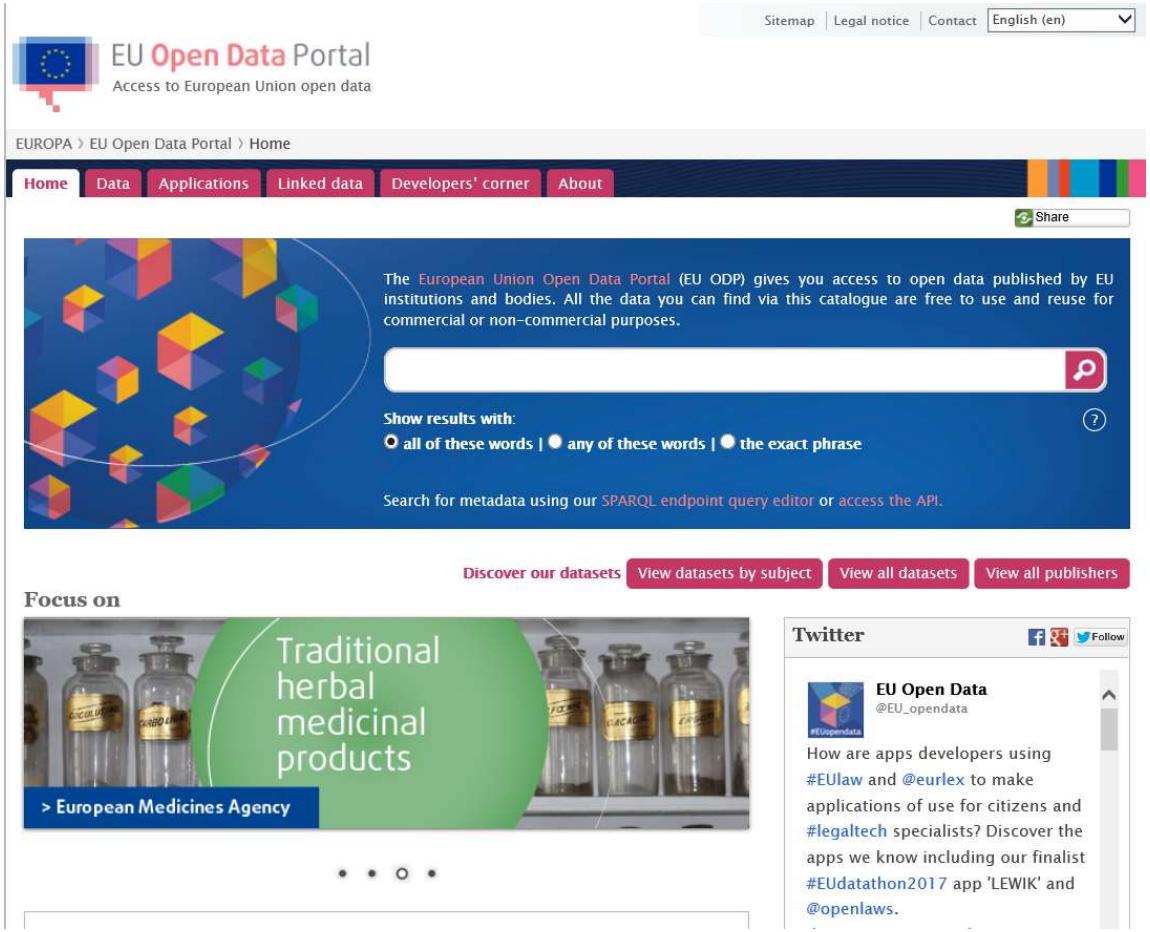
25 April 2018: EC Communication '**Towards a common European data space**' *(data package III)*:

- Review of PSI directive
- recommendations on scientific data
- guidance to private sector



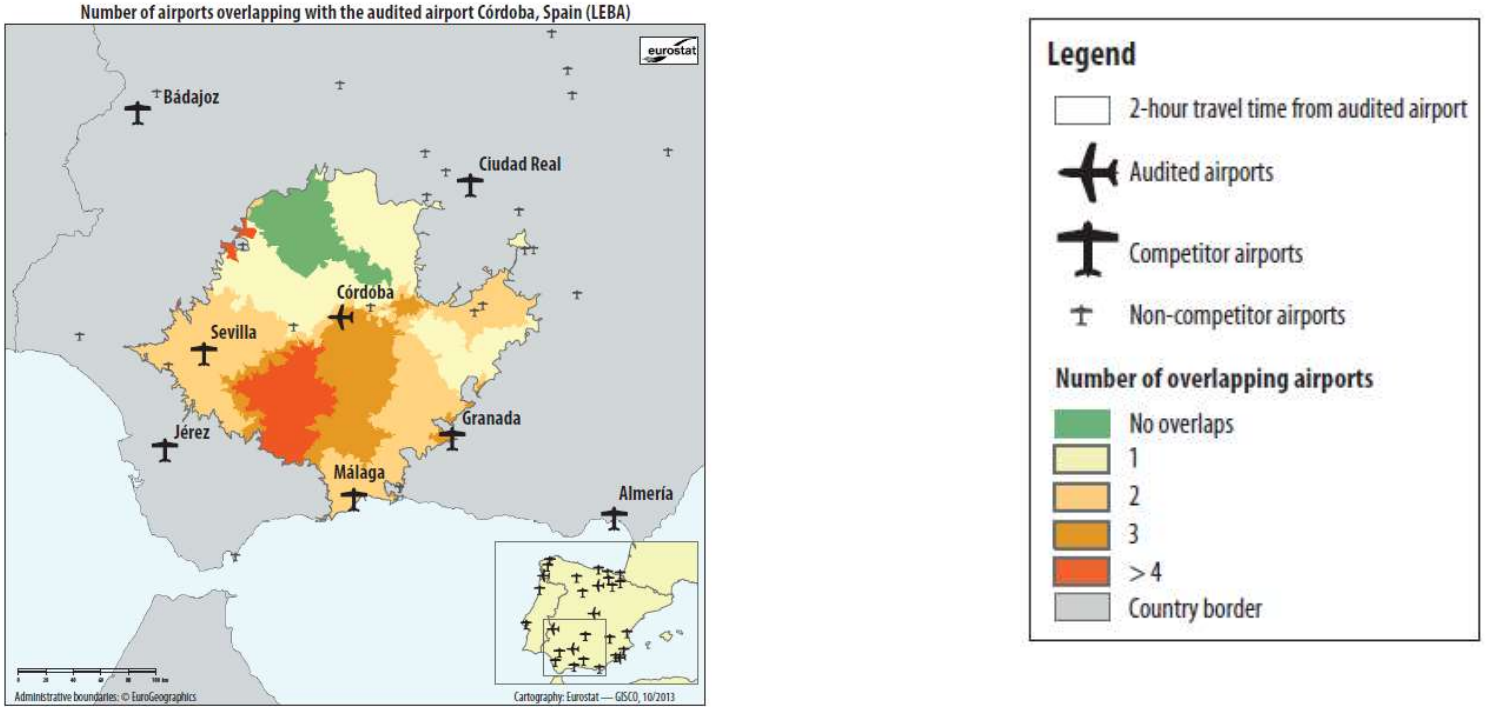
Open data

- EU ODP gives access to open data published by EU institutions and bodies.
- All this data is free to use for commercial or non-commercial
- Incudes also information on how to use open data
- For example a list of close more than 70 data visualisation tools (of which 15 have been developed by EU institutions!) as well as a knowledge centre with tutorials on how to use some of these tools.

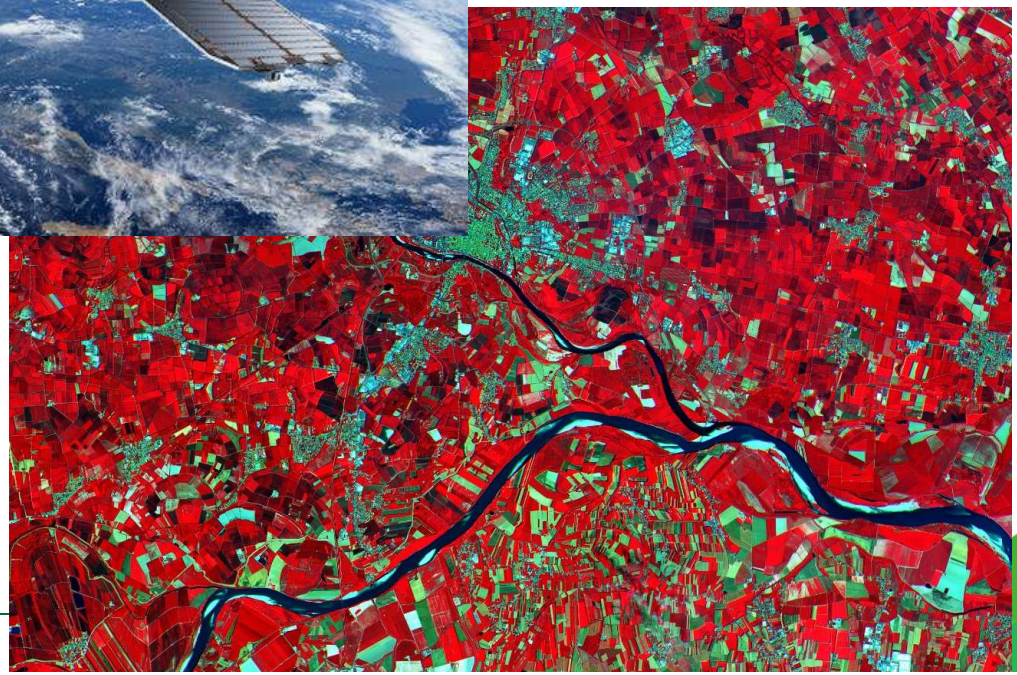


Member States

ECA Special Report No 21/2014: EU-funded airport infrastructures



Sentinel data

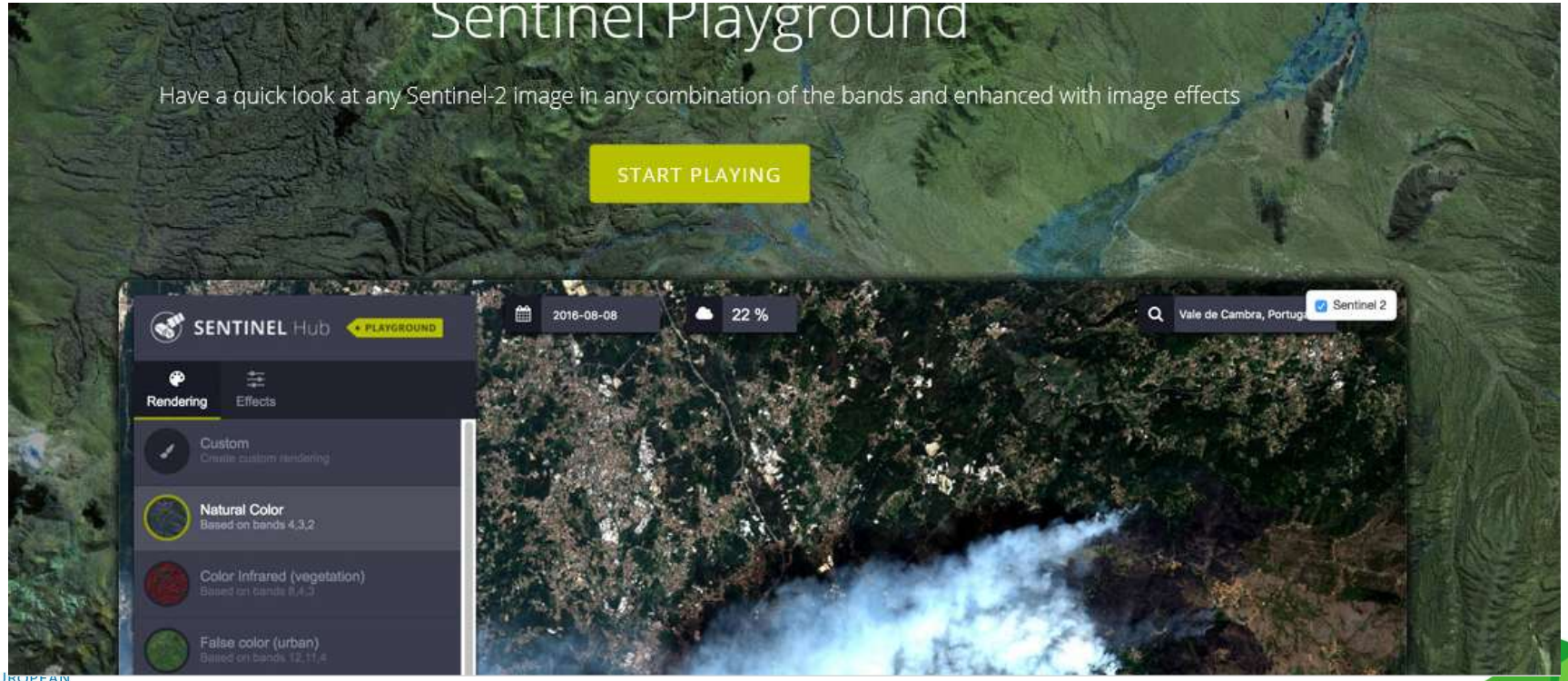


Sentinel data

Examples for applications include:

- Monitoring land cover change for **environmental monitoring**
- Agricultural applications, such as crop monitoring and management to help **food security**
- Detailed vegetation and **forest monitoring** and parameter generation (e.g. leaf area index, chlorophyll concentration, carbon mass estimations)
- Observation of **coastal zones** (marine environmental monitoring, coastal zone mapping)
- **Inland water** monitoring
- Glacier monitoring, ice extent mapping, snow cover monitoring
- Flood mapping & management (risk analysis, loss assessment, disaster management during floods)

Sentinel

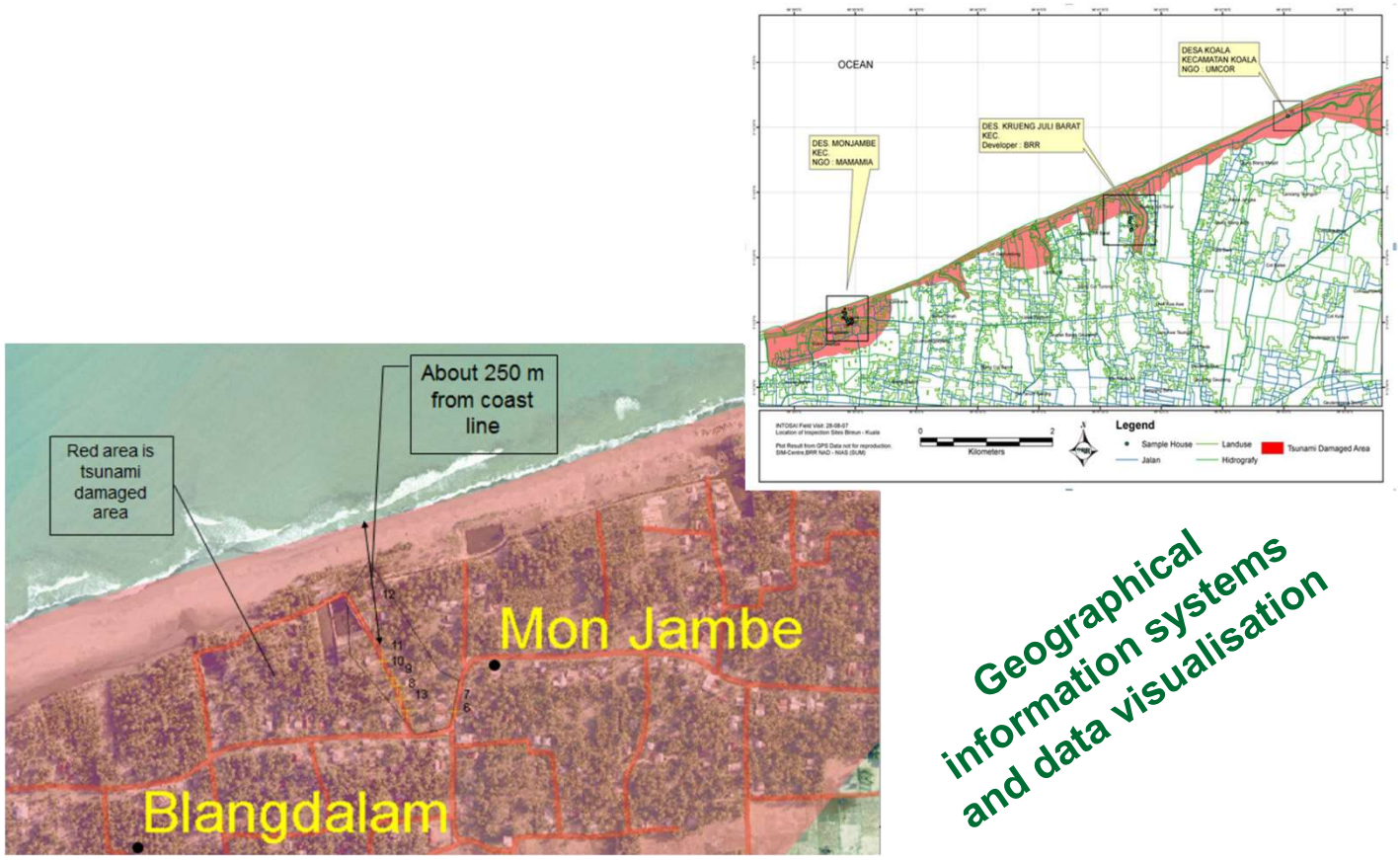

Sentinel Playground
Have a quick look at any Sentinel-2 image in any combination of the bands and enhanced with image effects
START PLAYING
SENTINEL Hub
PLAYGROUND
Rendering
Effects
Custom
Natural Color
Based on bands 4,3,2
Color Infrared (vegetation)
False color (urban)
2016-08-08
22 %
Sentinel 2

Tools

Tools - classification

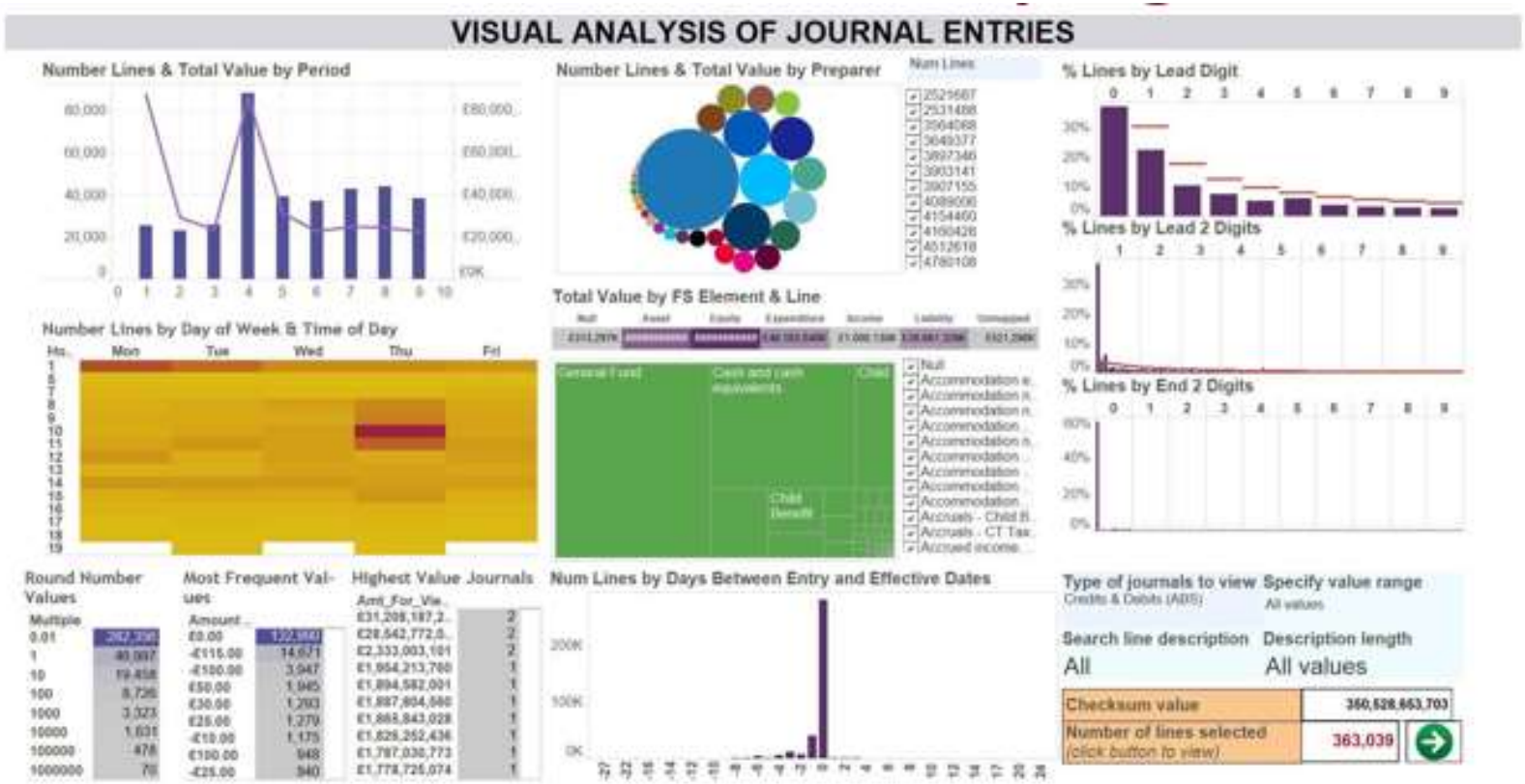
Office Automation	Specialised	Audit specific	Corporate	Open source languages	Artificial intelligence
(Excel)	(SPSS, SAS, Tableau, GIS)	(ACL, IDEA)	(SAP, Oracle)	(R, Phyton)	(Watson, Cortana)
In most of the organization these are the only used tools	Allow for an vast range of tasks from statistical analysis to data visualization.	Allow for advanced statistical analysis and require specialization via training	Used in audit, they require programming sckills	They allow for statistical analysis and text mining	They process natural language and machine learning

INTOSAI – Tsunami 2007



Geographical information systems and data visualisation

NAO example of Journal



Manual Journals present a number of risks. In a large population, it can be hard to see patterns. This dashboard allows a user to quickly identify unusual patterns, and investigate the individual underlying transactions. It also assigns transactions a risk score, to aid in their assessment.

Sentiment Analysis – executive summary of SR

Food waste

Water waste

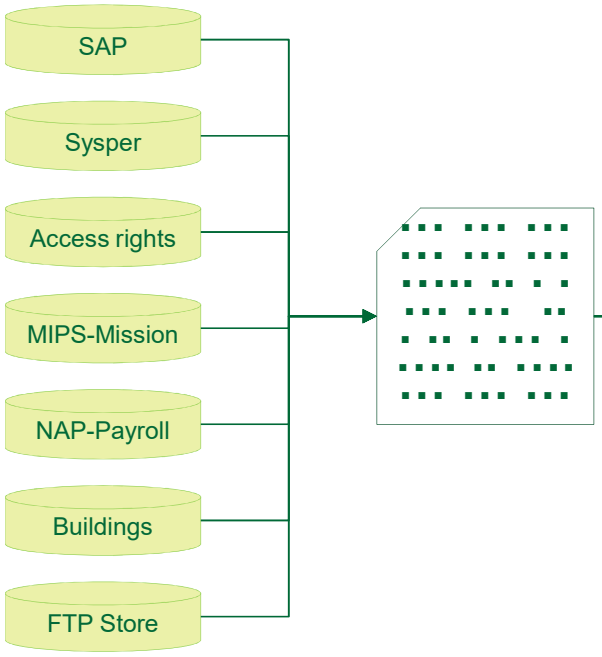
PWC auditing ECA IT

European Court of Auditors

IT procedures and related results for 2017

Access to systems and data	Change & Project Management	IT Operations	SAP configuration
• Logical access • Physical access • Segregation of Duties	• Governance • User Testing • Segregation of Duties	• Data retention • Business Continuity • Disaster Recovery	• Security parameters • Sensitive profiles • Sensitive programs
Outsourcing Management	**Salary grid configuration**	**Mission/MIPS Migration**	**Journal Entry review**
• Contracts • Committees • KPI reporting	• Automated controls • Data accuracy • Change Log	• Governance • Understanding • Logical access	• CAATs • 100% analysis • Analytics who booked entries and when

PWC auditing ECA



- Users having access rights not being there any longer
- Segregation of duties: IT staff having access as end user and as super-user
- Staff having access to critical infrastructure
- Super users with too many rights
- Critical systems with password policies not being updated

PWC Findings reporting

Short and to the point
100 % data analysed
Give value

Access to FTP Store (Medium risk)

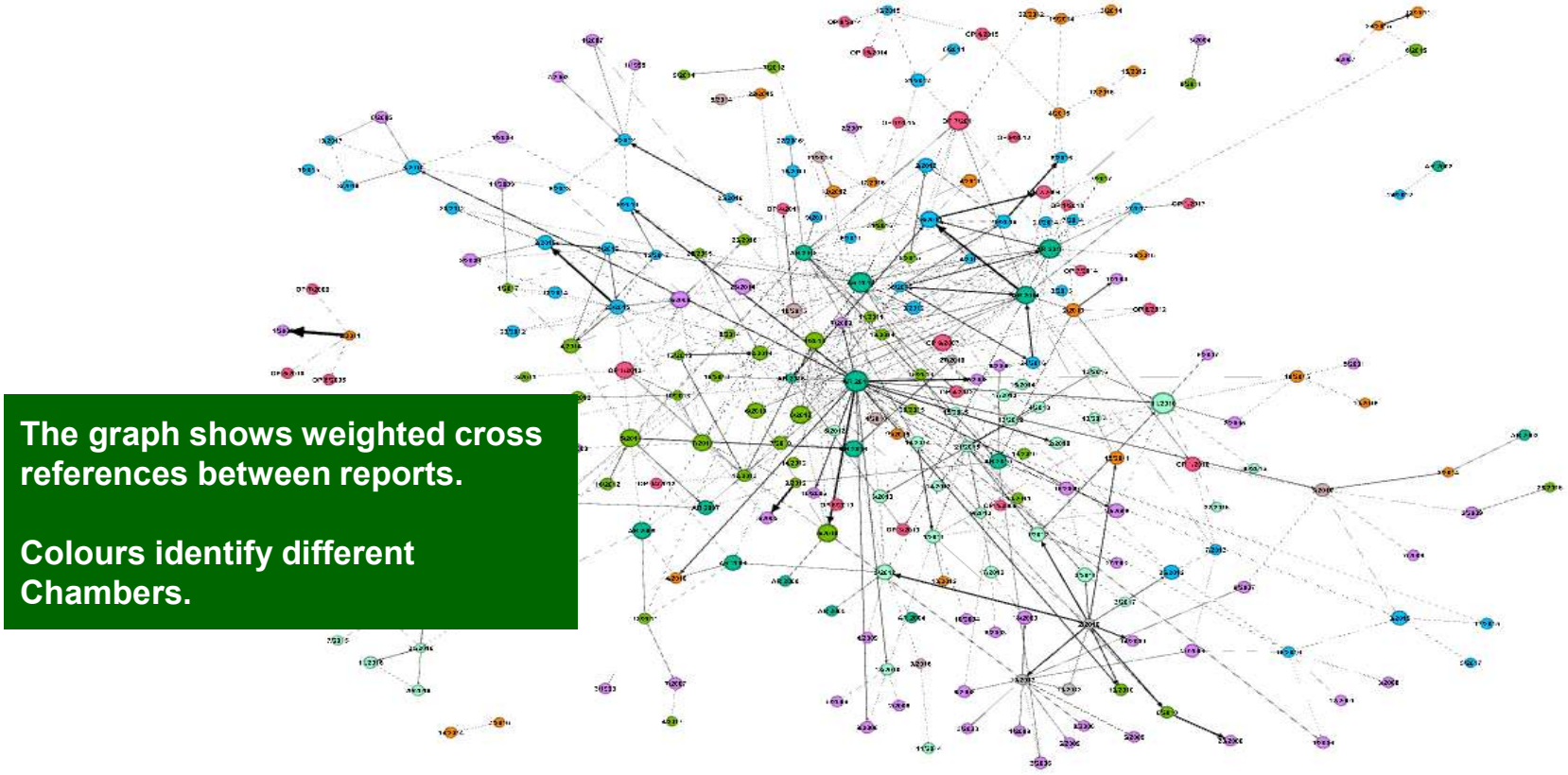
In order to be able to access FTP Store, users' IP addresses have to be whitelisted in the firewall (i.e. Access will not be blocked by the firewall). For 3 users from IT department and 1 user who left in 2015, the IP addresses were whitelisted while access to FTP Store is not required as per their job responsibilities. The whitelisted IP addresses have been already removed during the audit. We recommend to review whitelisted IP addresses on an annual basis.

Privileged access (Medium risk)

The Active Directory (AD) domain administrator accounts follow the default password policy (which enforces passwords to be changed once a year). However, the acceptable use policy of the Court requires that passwords of administrator accounts are changed at least every 90 days or completed by a second factor authentication. The AD parameter has been already adapted to follow ECA's acceptable use policy for administrator accounts.

Network analysis

Interconnection graph of SR, ARs and Ops



The graph shows weighted cross references between reports.

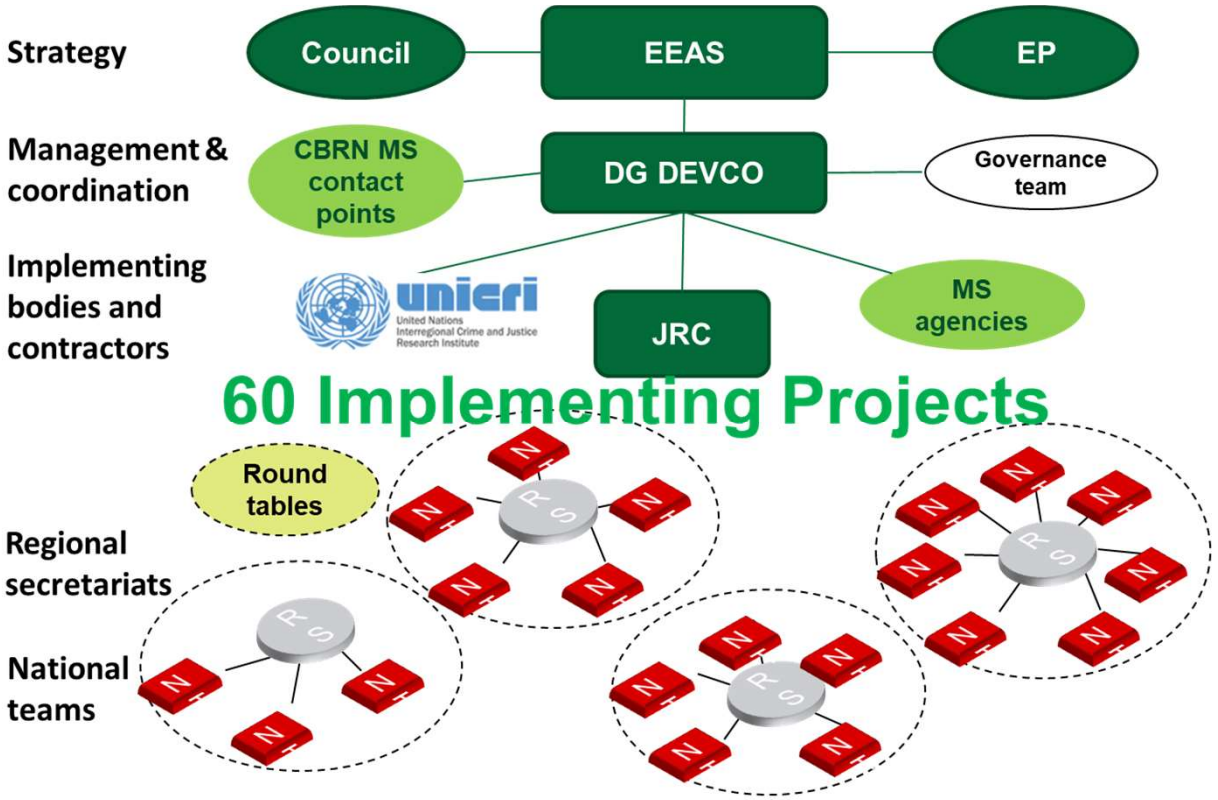
Colours identify different Chambers.

Exploring a network: EU Centres of Excellence (CBRN)

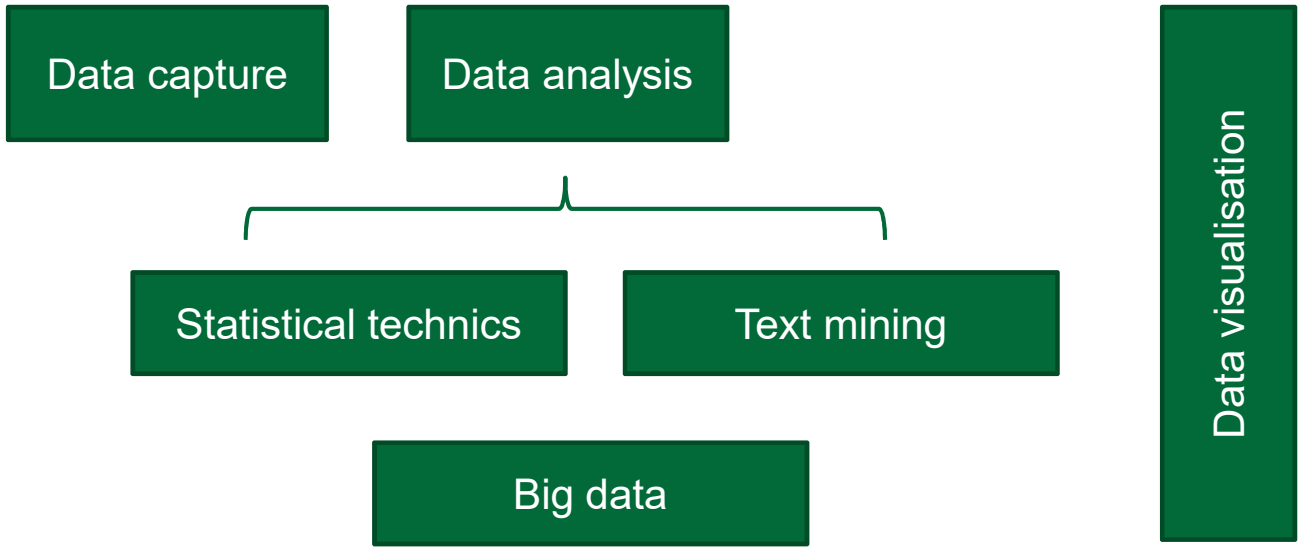
Case: Chamber III - Follow-up CBRN CoE (17CH3012)

Goals:

- model the EU Chemical, Biological, Radiological and Nuclear Centres of Excellence network
- measure effectiveness
- observe evolution over time



Techniques

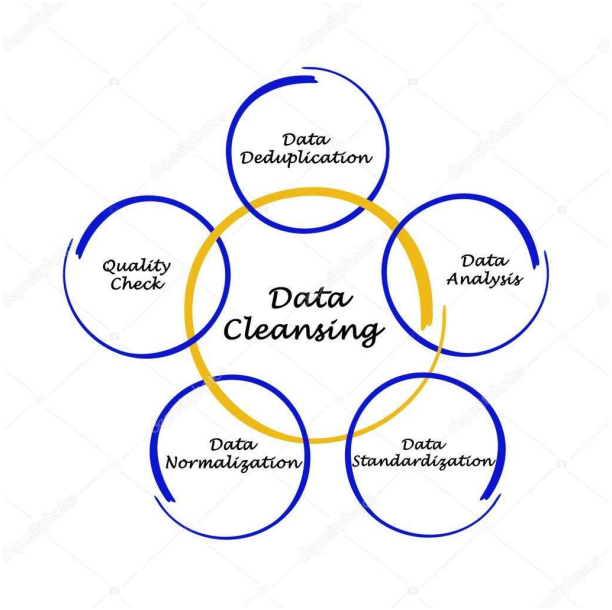

Data capture
Data analysis
Statistical technics
Text mining
Big data
Data visualisation

Data capture

- Regular data transfer
- Access to the auditee systems
- Data exchange via extranet sites (portal)

Data **preparation**

- Data extraction, transformation and load
  - Data cleansing



Data analysis (from data to knowledge)

Statistical techniques

Unknown model

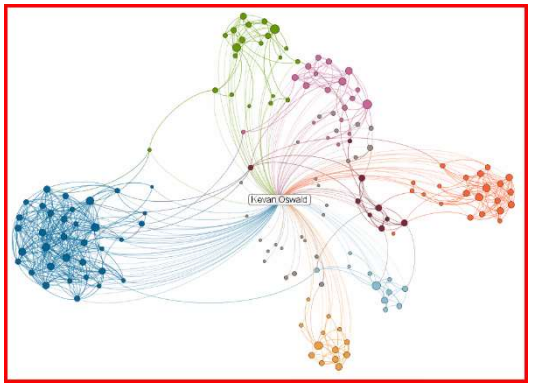
Exploratory what the data tell us?

- Visual. Allow for patterns and trends. Use of GIS (Geographical Information Systems)
- Data mining:
  - Explore the full dataset
  - Data will tell what is next
  - Analysis of outliers
  - Expert systems
  - Requires IT, statistics and business experts and skills

Known model

Analytical. Reply to specific questions: Is there relation between A and B?

Regression models, time series analysis, factor analysis, discriminant analysis, causality analysis,etc.



Big data

The Internet of Things and big data

Everyday objects have sensors , become smart and "talk " to each other.
Everything can be registered, measured and captured in digital format.

What is digitally captured ? what becomes data ?

What is big data?

- Massive amounts of data
- Combination of data from different sources
- Very complex data
- Generated by automatic processes (machine, generally)
- Requiring special treatment

Text mining

The goal is to turn text into data for analysis via application of **natural language processing** and analytical methods.

Text analytics linguistic, statistical and machine learning techniques that model and structure the information content on textual sources

- Semantic relations
- Sentiment analysis
- Pattern recognition
- Named entity recognition
- Speach tagging
- Machine translation
- …

…

Natural language processing

What Artificial Intelligence can do today in the field of document understanding

Summarise

Group by criteria

Search by meaning

Correlate different sources

What technology can do, ECB - 1

EUROPEAN CENTRAL BANK
EUROSYSTEM

DGC Senior Management Presentation
Frankfurt Am Main, 13 February 2018

DISC in Motion…
DG-IS DISC Team

Unstructured Data Processing on DISC - Document summarization example

1. Input to the Natural Language Processing Algorithm (~500 words)

Big data encompasses a wide range of highly complex information sources, characterized by granul (observed at the individual or product level), high frequency (such as intra-daily time series), lacl structure (data retrieved from text or internet sources), or combinations of the above.

The increased availability of big data creates a novel opportunity for central banks to improve on fulfilment of their mandate. In order to fully leverage the potential of ever-growing volumes of informat the European Central Bank (ECB) is compelled to develop the capacity to collect and store this data, well as analyse it by means of modern machine learning techniques. This encompasses a rang algorithms trained for data-driven predictive modelling.

Accurately assessing the present and forecasting the future are core activities of the ECB, continuously perfected by means of advanced analytical methods. The adoption of big data and machine learning techniques would represent an expansion of the existing statistical toolbox, further promoting informed decision-making. A range of novel economic studies illustrate how analytical tools designed for big data enable the construction of more precise statistical indicators, the detection of anomalies in economic patterns and the estimation of more accurate predictive models. Aware of these benefits, ECB business area experts have already experimented with machine learning techniques in their work, thus placing the bank at the forefront of cutting-edge economic analysis.

Alongside the conduct of monetary policy, the increasingly important role of the ECB as a financial market supervisor has entailed the production and analysis of more granular data. Given its pan-European access to a large and growing number of confidential datasets, the ECB enjoys a unique position to become a leader in the use of big data for monetary policy and financial supervision.

Big datasets used by the ECB require ample storage space and working memory, often exceeding the capacity of a single computer. Moreover, the confidential nature of the information managed and produced by the ECB prompted the need for a secure, in-house solution for data management. In order to enable ECB experts to harness the potential of available big data, DGIS has been working on improving the available IT infrastructure. To this end, DGIS put in place the Data Intelligence Service Centre (DISC) platform, a secure distributed computing environment for the storage of all datasets used within the bank, which enables general data processing capabilities and facilitates information sharing across business areas.

Over time, the techniques used by ECB experts have become progressively more demanding from a computational point of view. As these methods require the use of more processing power, analytical work conducted on a single laptop can become very cumbersome, slow or even unfeasible.

The necessary processing power for state-of-the-art analytics can be guaranteed by a system of computers that work together in a cluster. To this end, we will further leverage the Hadoop cluster computing system underlying the DISC platform. The parallelization of machine learning computations in the cluster would expand the possibilities for analytical work along two important dimensions: the use of large datasets and the deployment of algorithms with high computational demands, moving the ECB into the realm of big data.

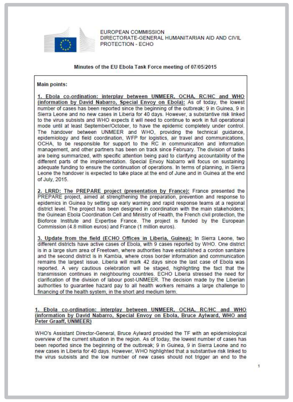
What technology can do, ECB - 2

3. Summary of the Summary Generated by the Natural Language Processing Algorithm (~50 words)

In order to fully leverage the potential of ever-growing volumes of information, the European Central Bank (ECB) is compelled to develop the capacity to collect and store this data, as well as analyse it by means of modern machine learning techniques.

Natural Language Processing

Document summarisation



Summary, 2% ratio
“The situation needs to be monitored closely in Guinea before we can announce a definitive slowdown of the epidemic: no new cases of primary contamination; all current and expected new cases are family or health

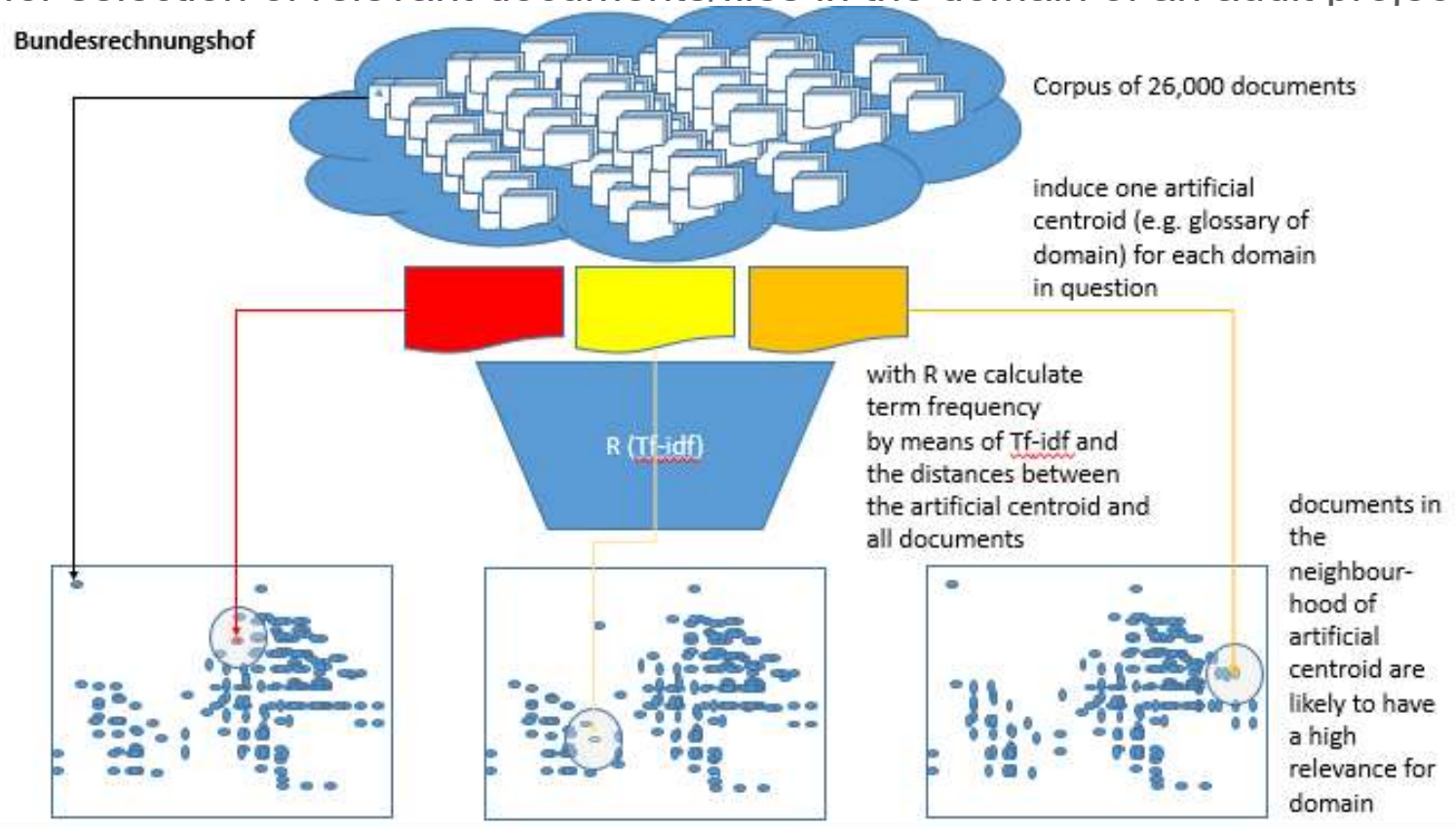
Most positive paragraph
“The response from national authorities, international organisations and partners show a clear improvement in the response”

Most negative paragraph
“The EU DEL Sierra Leone mentioned that there are no cases confirmed (to the contrary of previous messages from WHO). One suspected case was contaminated in Guinea and died around 12 March in SL. Contacts were isolated but they have tested negative.”

Case: Chamber III – Union Civil Protection Mechanism (14PAN253)
Goal: help auditors to quickly read through many documents and select the meaningful ones
ECALabers Emanuele Fossati

German SAI. Text mining methodology

Text mining for selection of relevant documents/files in the domain of an audit project.



Semantic search/navigation

Search by meaning

Report: SR/08/2016

Rail freight transport in the EU: still not on the right track

Recommendation

Reply (European Commission):

The Commission accepts the recommendation. Stakeholders have taken the lead in harmonising in certain areas (e.g. the common Framework for capacity allocation). An evaluation of the RFC regulation will be conducted in 2016 taking into account what has already been done or is in the process of being done, and assessing whether there are areas where harmonisation is needed through an intervention from the Commission.

Recognised Entities:

['The Commission (ORG)']

['Commission (ORG)', 'Framework (ORG)', 'RFC (ORG)']

Report: SR/08/2016

Rail freight transport in the EU: still not on the right track

Paragraph: 98 - Conclusion

Reply (EC/MS):

In addition to improving the regulatory and strategic framework, enhancing rail freight transport competitiveness requires a rail network adapted to specific freight needs, which entails making the best possible use of the available funding.

Under the 2014-2020 MFF, the newly creat almost entirely on projects with clear Euro cross-border rail projects and the ERTMS border sections and interoperability will ne freight services, since rail freight is particu and long distances. The Commission consi will continue to be made, to target rail inve TEN-T Programme.

Recognised Entities:

[]

['MFF (ORG)', 'CEF (ORG)', 'ERTMS (PERSON)']

The searched keyword “Train” is nowhere here, only similar concepts

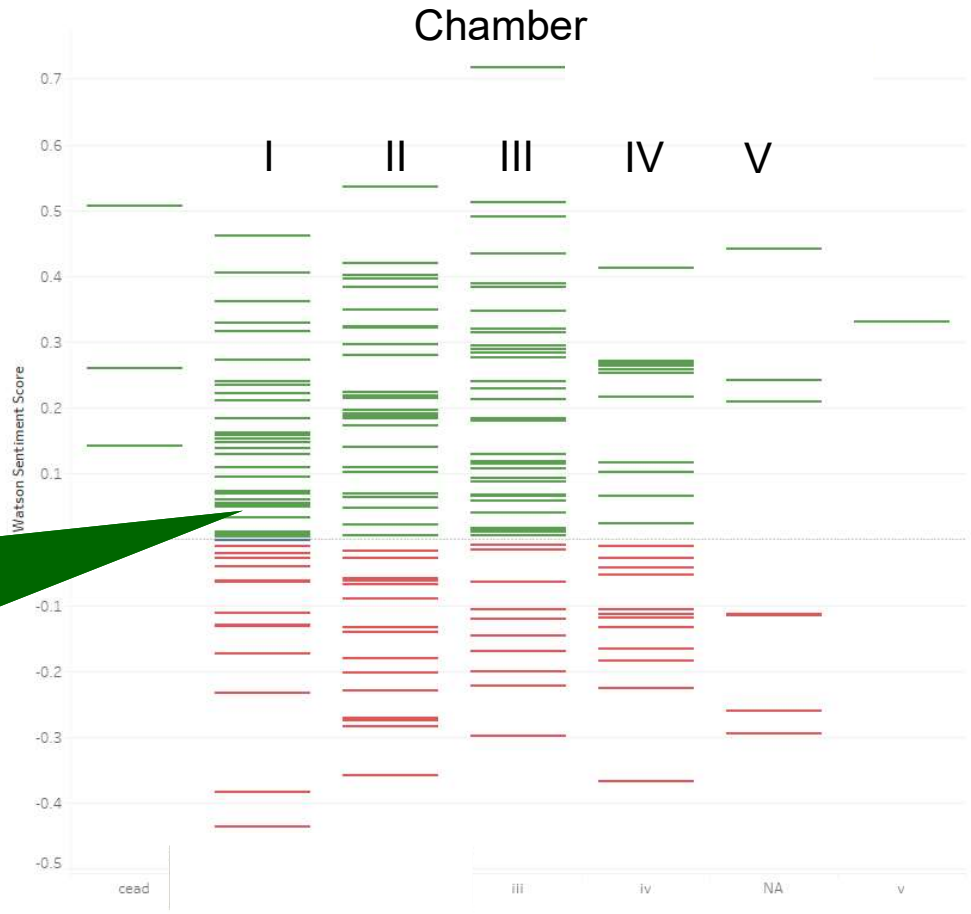
Important Entities are recognised!

Sentiment analysis

Sentiment scores of Special Reports

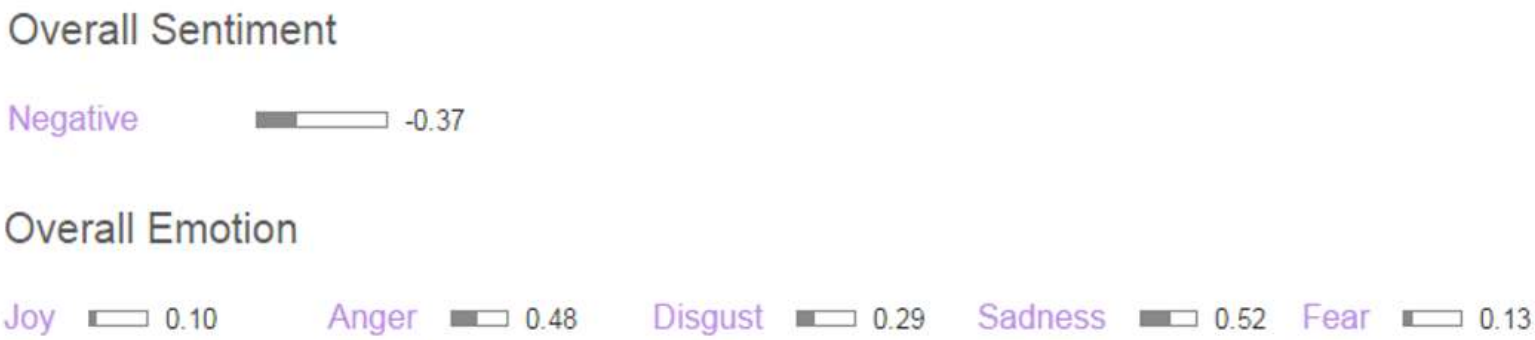
How "positive" or "negative" is the wording used in a document

Overall balanced distribution.
CH I is slightly less positive CH III is slightly more positive
In wording.



Sentiment Analysis – executive summary of SR

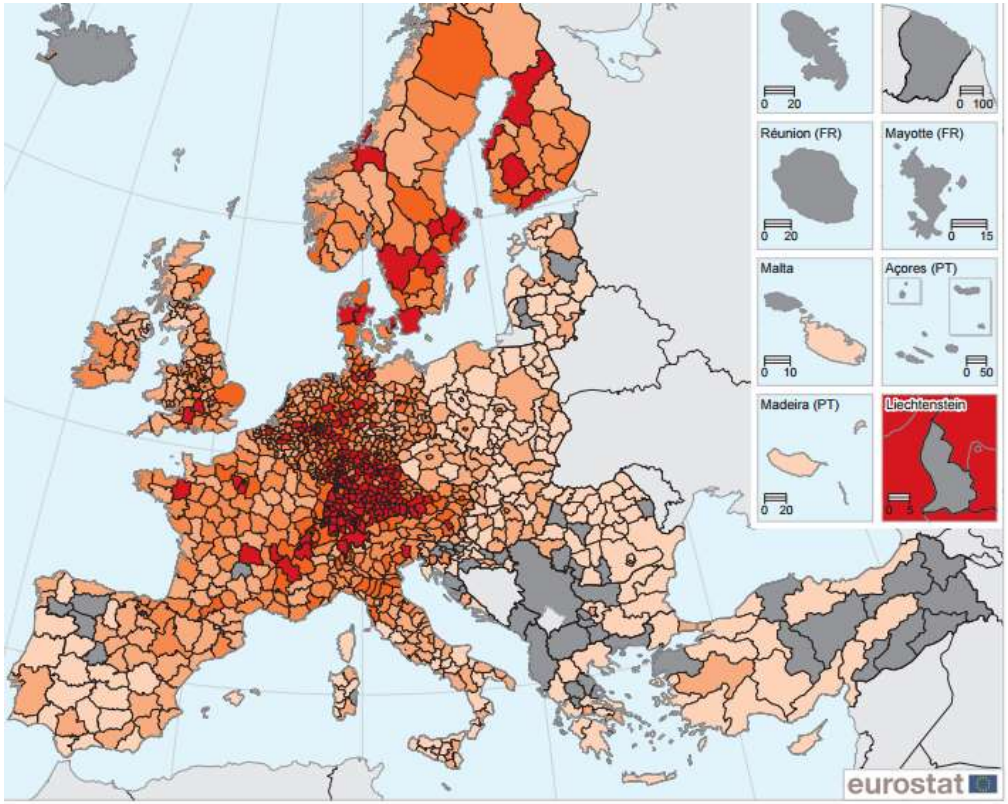
Food waste



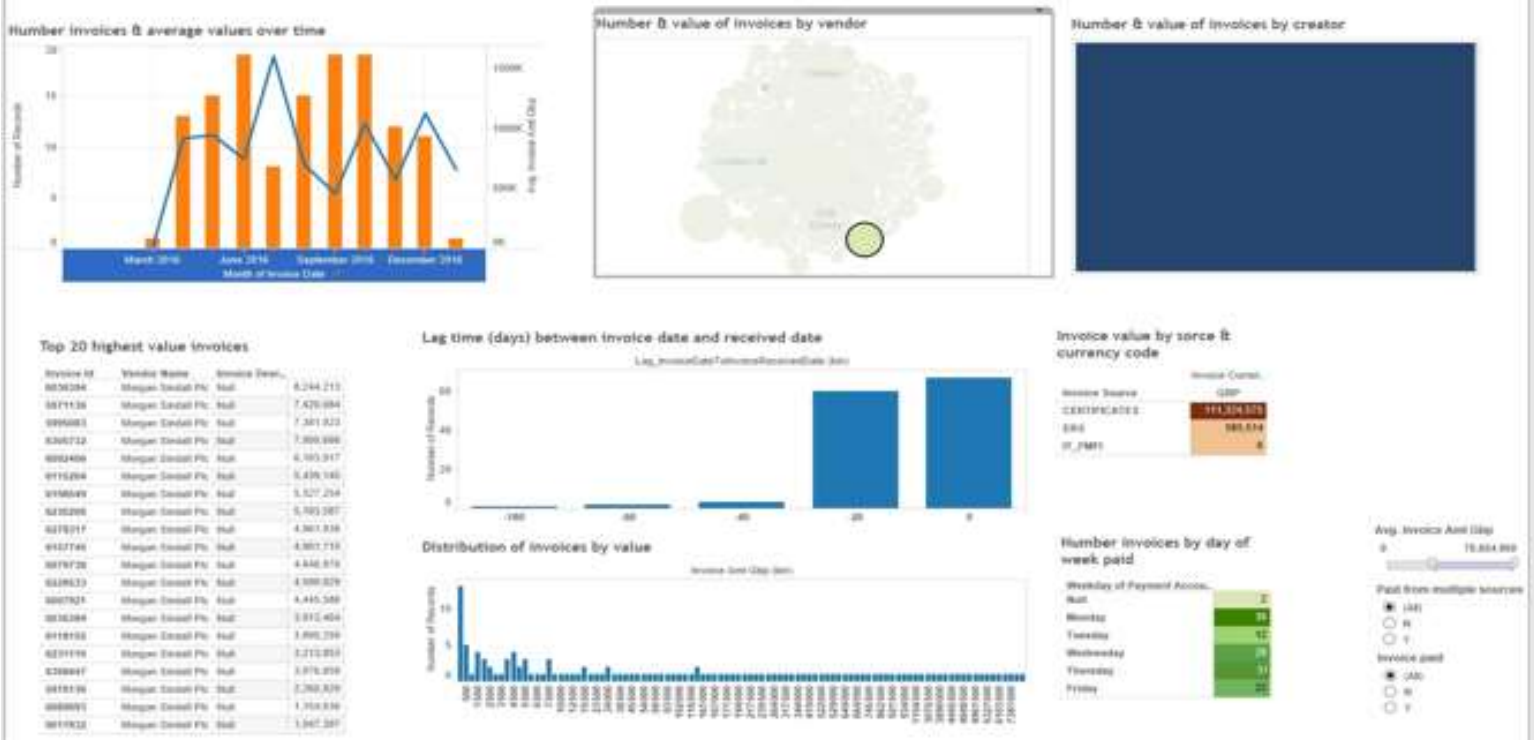
Water waste

Data visualisation

- Tabular
- Graphical
- Word clouds
- Infographics
- Dash boarding



Purchase to pay (P2P) analytic – invoices analysis

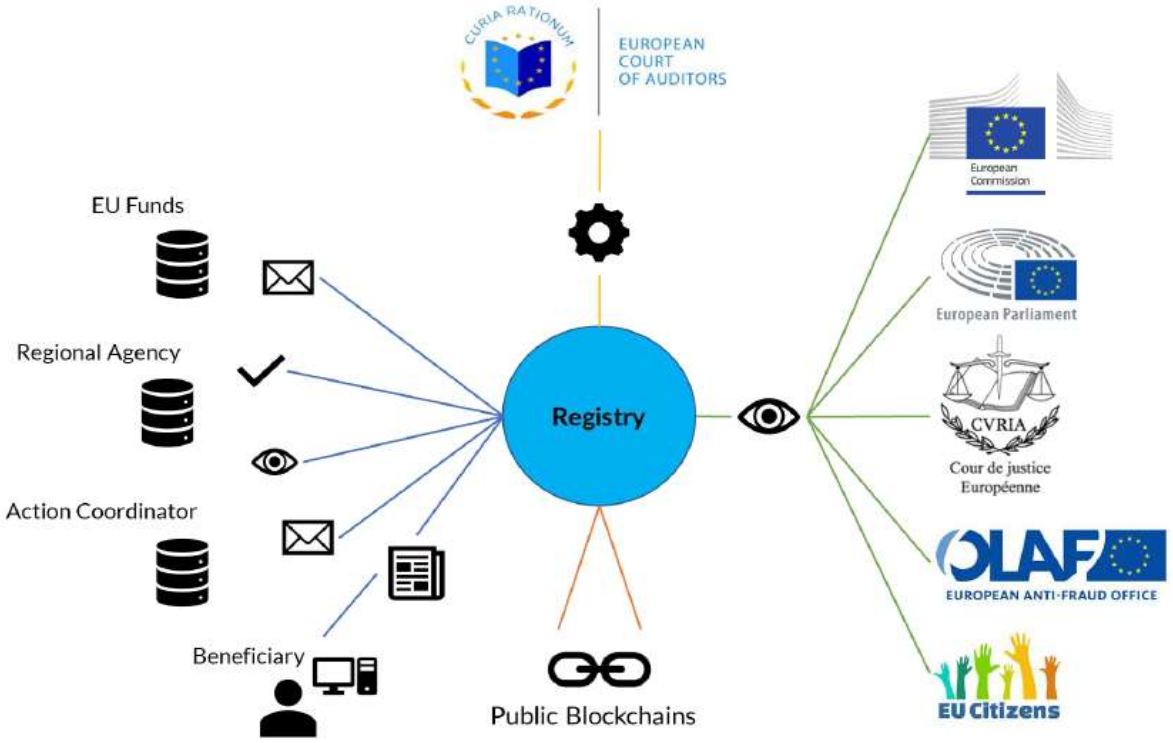


Building a series of dashboards that allow audit teams to consider their knowledge of the business and identify areas of interest is a key aim of our P2P work.

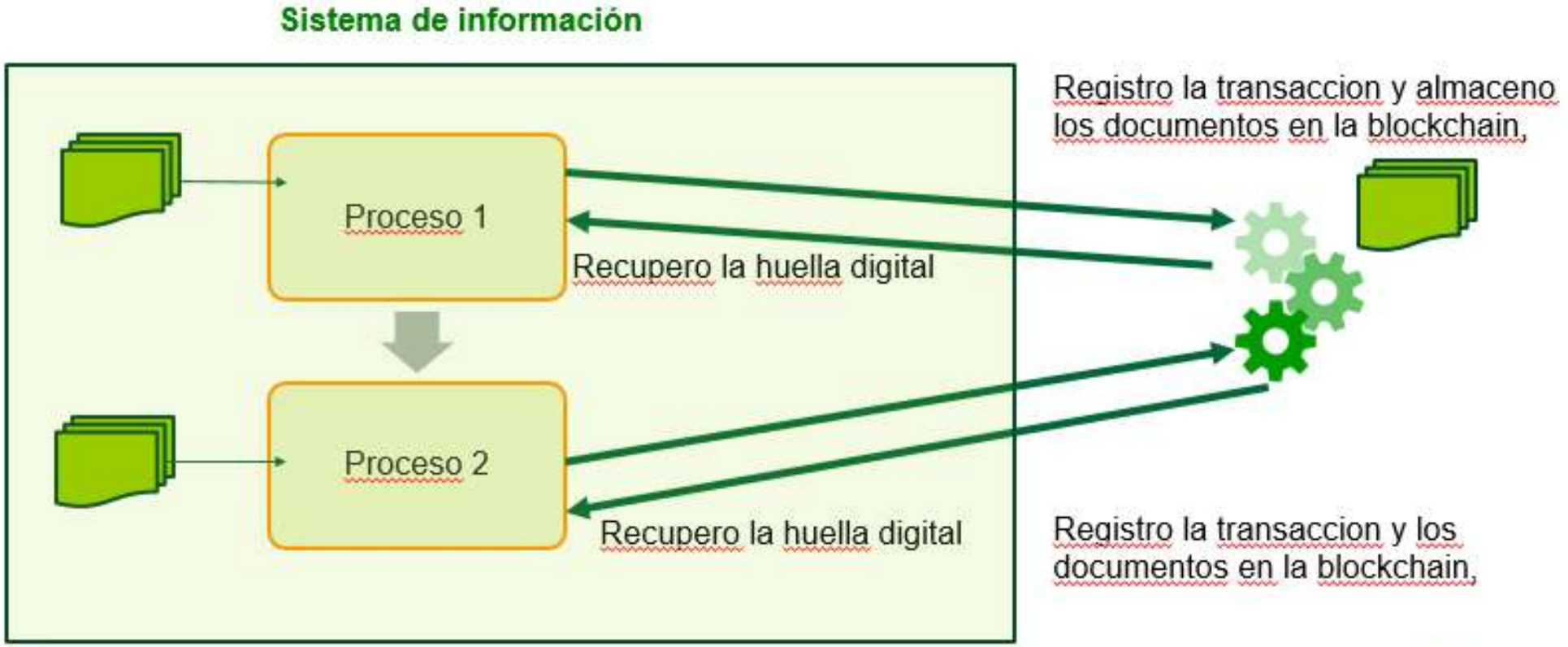
NAO experience of applying data analytics to performance and financial audit:
24th UN/INTOSAI Symposium 2017

National Audit Office

Blockchain. Contol by design.



Diseñando sistemas de información. El “control” por diseño



Process Mining

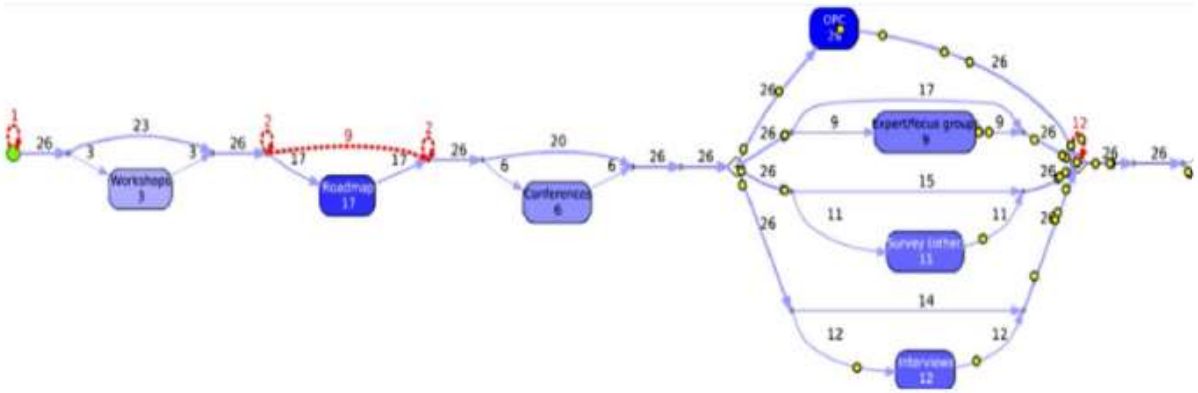


Figure 4: ProM Lite Process Model. Source: ECALab

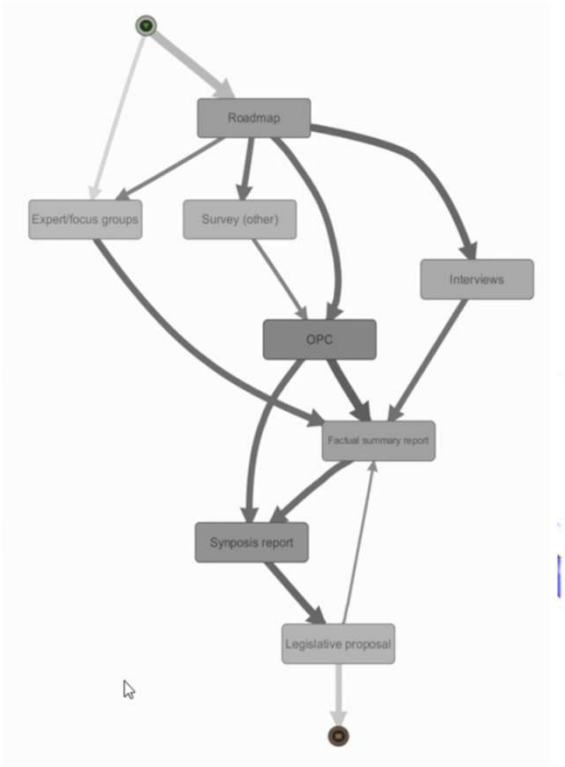
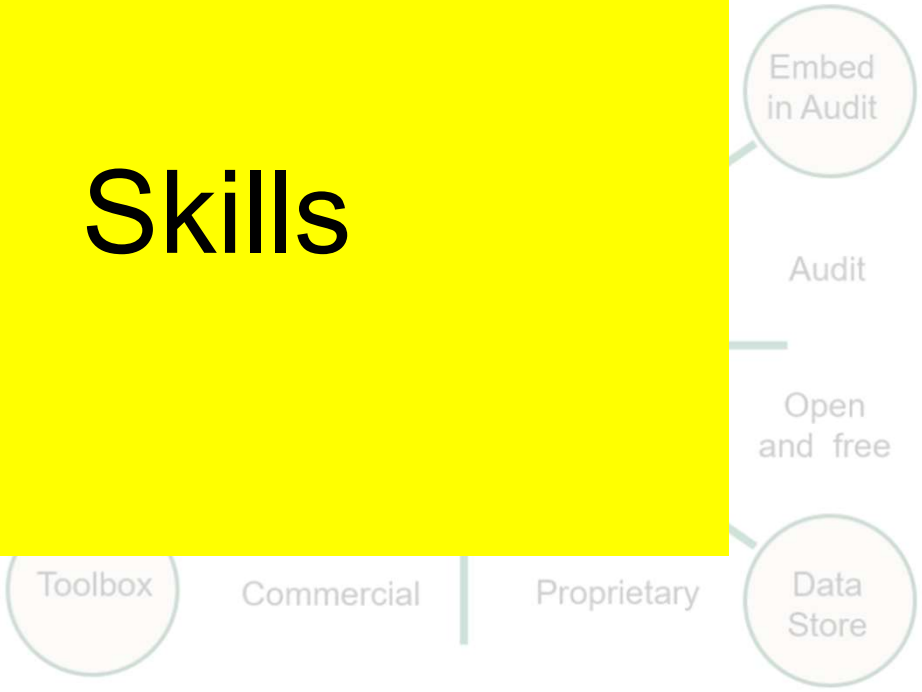


Figure 5: Fluxicon Disco Process Model. Source: ECALab

Skills


Embed in Audit
Audit
Open and free
Data Store
Proprietary
Commercial
Toolbox

Prepare the auditors of the XXI century

New areas of knowledge:

- Information technology
- Data science and statistics
- Communication
- Modelisation

The new auditor:

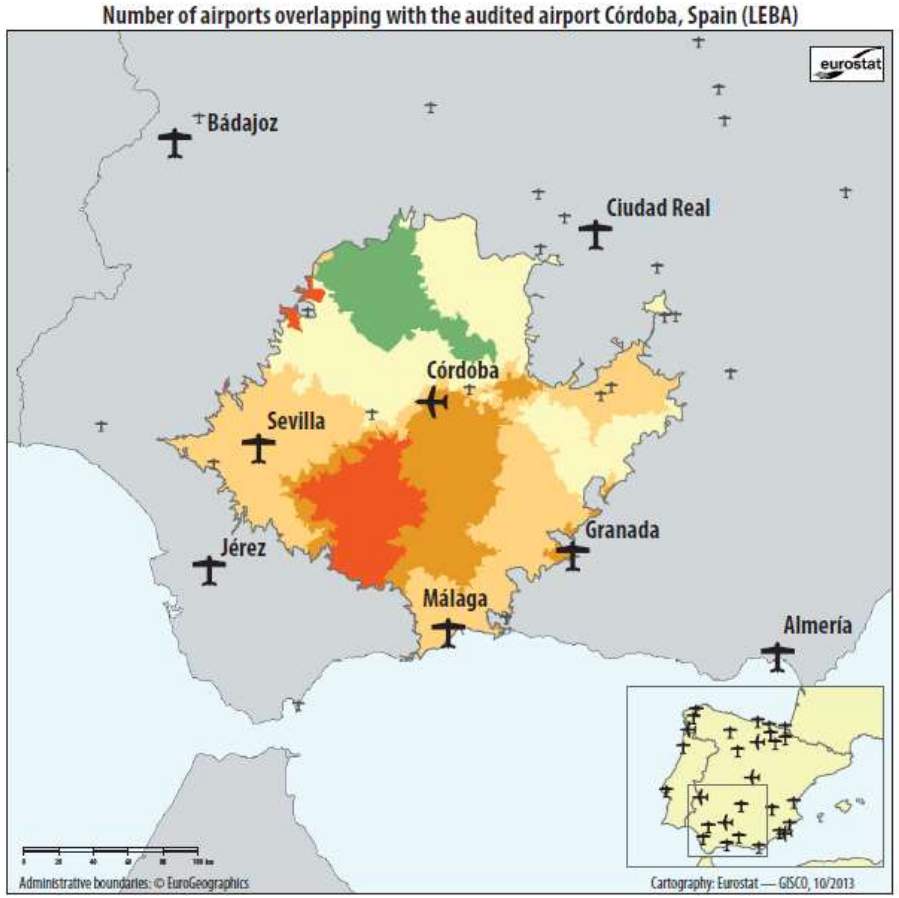
- must feel confortable on these domains and
- must be able to work on teams where different expertises are combined: **cross-functional team**.

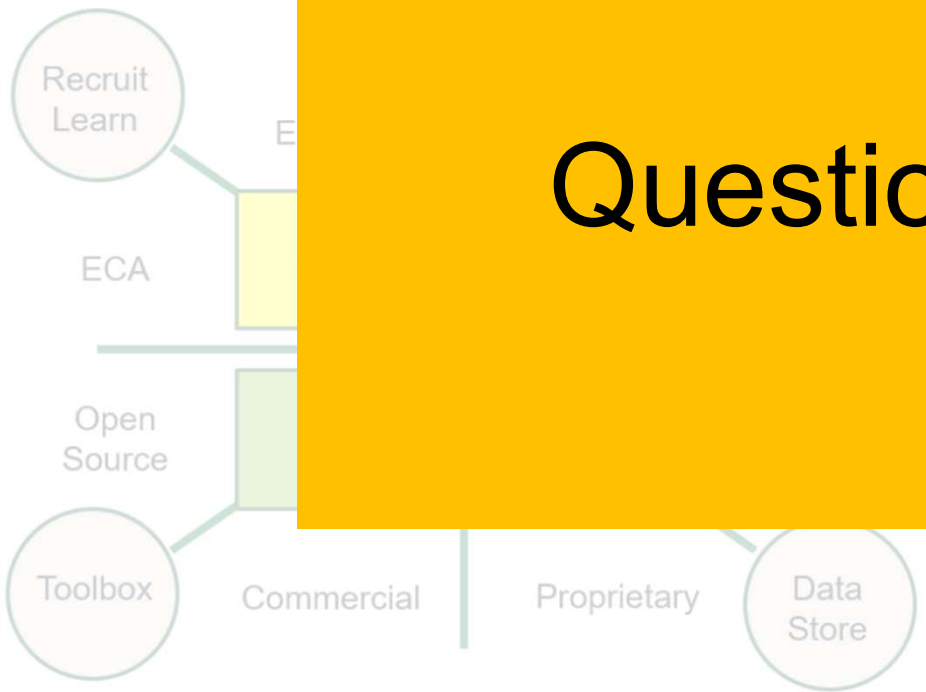
Skills required – cross-functional team

Auditor	• Understands audit objective • Identifies opportunities for data analysis • Generates the demand • Solves audit problems • Asks questions
Data scientist	• Expert in analytical tools • Data extractions and manipulations • Statistical analysis knowledgeable
Data management expert	• Analyses and programs algorithms • Administers the database • Expert in data extraction, cleansing and transformation
Domain expert	• Knows the domain

How to acquire the skills

- **Recruitment**
- **Training**
- **Collaboration agreements**
  - Other audit institutions: Eurostat,
  - Academia: Lorraine, Pisa




Recruit
Learn
ECA
Open
Source
Toolbox
Commercial
Proprietary
Data
Store

Questions

Questions

- From policy scan
- From the data analysis itself
- Not only give replies but also ask questions.

Benefits of the digitalisation

How de we imagine the near future

Continuous and predictive audit : *("…audit today the problems of today…")*

- Auditors have remote access to the auditee systems or receive frequent updates of the data.
- Algorithms are developed containing the rules for identifying irregularities and alerts.
- Audit is not "ex post only" anymore. Auditors can continuously interact with the auditee indicating potential irregularities

Risk analysis on full population: *("…with more confidence")*

- Auditors carry out risk analysis on full populations, using data and information available from different sources.

How de we imagine the near future

Process all available information: *("…audit all problems of today")*

- Auditors are assisted by technology in processing all information they can access in any digital form.
- They can navigate through masses of text, images and figures and interact with the information
- They also get suggestions/alerts for abnormal or atypical behaviour worth examining.

Audit automation: "*…with fewer resources and fewer errors*"

- Manual, recurrent processes related to audit as well as processes linked to delivering standard services are automated.
- Machines learning from the past, are used to leverage past experience and reuse past work to a maximum to gain insights from all information available, i.e. not reinventing the wheel.

How de we imagine the near future

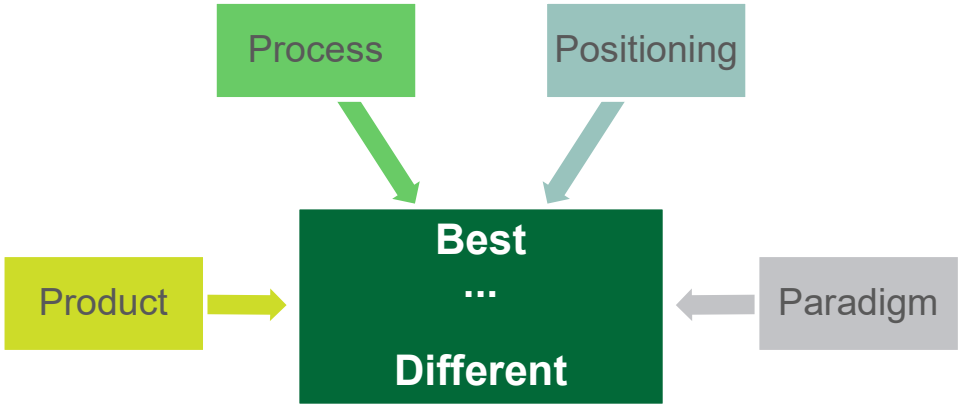
eGovernment: (…with more transparency")

- Self-controlled processes have been set up, building on the already implemented good principles of eGovernment, i.e. all is digital by default, all is done once only, information systems across the EU are interoperable, trustworthy, secure, open and transparent.
- Auditees and beneficiaries register in real time evidence of transactions in any digital form in secure, immutable and directly auditable ways.
- Auditors can access and audit this information at any time.

These are the actions mentioned in the Tallinn declaration on eGovernment adopted at the ministerial meeting during the Estonian Presidency of the Council of the EU on 6 October 2017

Innovation

Innovation: the “4 Ps” model



Tidd J., Bessant J. (2016). ”Managing innovation: Integrating Technological, Market and Organizational Change, 5th edition”. *Wiley.* Chichester, West Sussex.

Closing remarks

Prepare auditors for the XXI century

- **Strategic thinking is needed at the top level : p**romote innovation and experimentation in a "no-regret" and learn fashion
- **Data is an asset :** Full life-cycle, new roles and responsibilities
- **Driven by audit, not by technology**
- **Critical thinking and intellectual curiosity**
- **Boost collaboration and community of practice:** Learn from others
- **Close the skills gap:**invest in hiring the right people and build a pool of experts

Creating a network to share knowledge and expertise

ECA IT Master Plan 2018-2020

Goal1: Technology for audit

- Implementing the mechanisms to share knowledge with selected external stakeholders like experts from other EU institutions, SAIs, etc. active in data analysis work.

Goal 2: Connect

- Establish sharing platforms between ECA and SAIs in member states and the world, to share knowledge and insights on audit related issues.

Actions:
Create a “second internet” to publish the non official information
Organise a conference last quarter of 2019 in Luxembourg to launch the initiative.

The future…

- Impossible to predict impact of emerging technology.
- Transformation is going to be very rapid and disruptive.

We need to :

BE ADAPTABLE

- flexible approach when thinking about the future.
- diversify and improve our analytical tools to uncover audit opportunities.

BE PREPARED

- Build up in-house digital and analytic capability

HERE IS Edward Bear, coming downstairs now, bump, bump, bump, on the back of his head, behind Christopher Robin. It is, as far as he knows, the only way of coming downstairs, but sometimes he feels that there really is another way, if only he could stop bumping for a moment and think of it.

A.A.Milne. Winnie the Pooh