# 8th International Integrity eSeminar
# 25 February 2021

**Are emerging technologies helping win the fight against corruption?**

Dear all,
I warmly thank the SAI of Hungary for the kind invitation to participate as speaker in the 8th International Integrity eSeminar and I am very pleased to give our contribution.

Today, we deal with the potential of the Information and Communication Technology (ICT) tools on corruption.
Corruption continues to be one of the most pervasive anti-social vice in our societies. Often and especially in developing countries, corruption is associated with government officials and politicians who use it as the oil to lubricate the government machinery that guarantees their stay in power. Because corruption entails serious threats for economic, social, and political development; the world has, both at global and national levels, witnessed relentless efforts to fight it.

At global level, many cooperation initiatives have been undertaken such as the **United Nations Convention against Corruption** ratified by 178 countries and the European Union in December 2015; the **Convention on Combating Bribery of Foreign Public Officials in International Business Transactions** signed by the Organisation for Economic Co-operation and Development (OECD) which counts the 34 OECD states and seven other countries (Argentina, Brazil, Bulgaria, Columbia, Latvia, Russia and South Africa; several regional OECD initiatives which have developed ad hoc networks. The **Governance and Anti-Corruption Strategy** adopted by the World Bank (in 2007 and then updated in 2012) with the aim to tighten the monitoring of cash flows in all the bank's projects and loan programmes.
Anti-corruption measures are also part of the **G20's agenda**. In recent years, the G20 has played a critical role in global and national anti-corruption efforts. Indeed, the ACWG has led the G20 anti-corruption efforts coordinating the collective and national actions taken by its members.
Since December 1st, 2020, Italy holds the Presidency of the G20 and will have the opportunity to achieve such important goals laid down in the agenda.
Moreover, the European Union has adopted a range of anti-corruption treaties and protocols. In addition, the **European Anti-Fraud Office (OLAF)** also deals with corruption in the context of corporate crime.

The Council of Europe has founded the **Group of States against Corruption (GRECO)** which uses a peer-assessment process to determine the extent to which its member states respect and comply with the Council's anti-corruption standards.

While confirming the need of the most extensive and strongest international cooperation, nations have also implemented and strengthened their respective policy and regulatory frameworks to make their war against corruption more effective.

In Italy, we have adopted a solid anticorruption legislation designed to try and catch more shady and mimetic unlawful conducts and to strengthen integrity in the public sector at large. Pursuant to Law N. 190 in 2012, the National Anticorruption Authority (A.N.AC) has been established. Indeed, the ANAC is tasked with the analyses of the causes of corruption and with the drafting of the anticorruption strategy, through regulatory and supervising function.

Each public administration is required to adopt specific measures to prevent the occurrence of acts of corruption or bribery. These include the adoption of a three-year Anticorruption Plan, the appointment of a compliance officer, and the adoption of a Code of Conduct.

Furthermore, the Law N. 179/2017 has laid down special "Provisions for the protection of whistle-blowers who report offences or irregularities which have come to their attention in the context of a public or private employment relationship". Whistle-blowers will not suffer dismissal, sanctions, or discrimination for having reported misconducts. The whistle-blower's identity cannot be disclosed without express consent.

In this scenario, the *Corte dei conti* plays a leading role in deterring and combating public sector corruption. We carry out compliance, financial and performance audits in addition to judicial functions. The exercise of audit and judicial functions which are interconnected and complementary represents an added value for ensuring accountability and transparency. It is out of question that we are well positioned to make a great contribution to corruption fighting. In our case, it is recognized that the strength of the *Corte dei conti* in fighting and preventing corruption derives also from its Public Prosecutor Office: an essential tool for sound administration, with important preventive and deterrent effects, also in corruption cases.

Combating fraudsters is at the heart of SAIs and a persistent challenge.
In the Integrity Seminar context, our attention is focused on the use of new technologies and their potential to combat corruption.

Firstly, we must take into consideration that the world is becoming more and more digital. Today the reality is data. Data is being generated and collected all around us at every moment, for example, geographical information, statistics, weather data, research data, transport data, energy consumption data or health data. The increasing capacity to make sense of 'big data,' thanks to the advance of data analytics and management tools, is leading to technological innovation and the development of new services and products.

A new digital frontier known as Artificial Intelligence (AI) has emerged and fundamentally transformed every facet of our lives by making it easier and faster to access goods, services, and information regardless of geographical location. Digital channels are now being used in delivering more transparent public services and enable information to traverse through and across government structures while allowing the public at large to interface with government digital platforms that host vast volumes of information. This allows for greater transparency, accountability while limiting discretion vested in public officials –all key factors in minimizing opportunities for corruption.

The proliferation of data alongside advances in computational capacities have ushered in the golden age for Artificial Intelligence (AI), where algorithms and models can reveal anomalous patterns, behaviours, and relationships—with speed, at scale and in depth—that was not possible even a decade ago.

The advent of Artificial Intelligence with its peripheral technologies such as machine learning, robotics process automation and object recognition has expanded the cyber space in a manner not seen before.

Given the above, it is therefore important to establish if the new digital technologies can represent new opportunities to enhance the anti-corruption crusade.

Will ICT/Artificial Intelligence help societies emerge victorious in the war against corruption? And is the world going to witness the emergence of robots capable of replacing human intervention on both the demand and supply sides of corruption?

It is beyond doubt that public sector organizations are taking advantage of powerful algorithms to identify underlying patterns generally hidden from plain sight as well as more prospectively detect and address red flags before they manifest into significant issues. And we are all convinced that new technologies significantly help SAIs to perform oversight work using available resources to produce high-quality results. Indeed, this digital transformation did not leave the world of auditing untouched. On the contrary, auditors have already begun to embrace the digital revolution in order to execute their functions in the most efficient and effective way. The availability of non-traditional datasets, together with enhanced data analytics technologies,

represents a great opportunity for auditors in fulfilling their duties. For example, auditors can now more easily combine datasets to analyse a particular phenomenon. In doing so, they can better identify weaknesses, inefficiencies, and trends in order to make recommendations to auditees. Audit processes should continue to harness the latest technological developments in **data analytics**, **machine learning** and **artificial intelligence**, to allow auditors to perform their financial, compliance and performance audits in a faster and more innovative way, drawing conclusions that would have been impossible to reach via traditional statistical sampling. By automating data collection and processing, these tools could spare auditors from time-consuming manual data collection and analyses, thus allowing them to focus more directly on the analysis of insights to identify potential risks and fraud cases more easily. Finally, digital auditing – using digital means available to arrive at an audit opinion – is set to improve the **transparency and accountability** of public authorities towards the citizens they serve.

Importantly, while Artificial Intelligence tools identify patterns and exceptions for further investigation, they do not replace professional judgments of experienced auditors in detecting potentially fraudulent activities. Although AI can sift through large volumes of data with tremendous accuracy, human intelligence is still an essential element for determining context-specific, proportionate, and nuanced actions stemming from algorithmic outputs. This symbiotic relationship means Artificial Intelligence will assist Supreme Audit Institutions (SAIs) work and will change how that work is carried out—requiring different skills to harness AI's capacity to drive effectiveness and efficiencies.

Indeed, Artificial Intelligence algorithms do not understand the difference between fraudulent and non-fraudulent transactions. Instead, these algorithms identify anomalies, such as unusual transactions between accounts. Human subject matter experts are still needed to analyze these anomalies to determine whether potential fraud exists.

While Artificial Intelligence systems can indicate risk, auditors are needed to further investigate actual conditions, causes and effects. It is imperative Artificial Intelligence be employed as an assistive technology to augment the audit process and equally important that Supreme Audit Institutions (SAIs) cultivate skilled manpower to harness Artificial Intelligence technology.

With this in mind, the shift towards digital audit still requires a review of the working methods and organisation of audit institutions. The purpose of the audit function has not changed: with its focus on accountability and transparency, it aims to foster stakeholder trust concerning the efficient, effective, and economic use of resources. The methodology on how this is to be

achieved and the related regulatory framework should be reviewed to ensure they correspond to the digital reality.

In such a scenario, we must move forward modern Supreme Audit Institutions. Building on what we have already put in place, and what remains to be done. We need to experiment with new methods in innovation labs. We need to identify best practices and pointing out the bottlenecks hindering the usage of data analytics. And we need to learn from each other through cooperation and exchanges.

Now, I would like to share with you our experience related to the digital transformation in Italy, as well as an interesting example of predictive Justice project implemented by the Scuola Superiore di Sant'Anna di Pisa (one of the most famous and prestigious University in Italy).

We all know that **digital transformation of the Public Administrations** is essential in order to create new opportunities for growth, simplify bureaucracy and make policies more transparent and effective.

The digital transformation strategy of the Public Administration (PA) is contained in the **Three-Year Plan (**annually updated). On August 2020, the new strategic policy document that will accompany the digital transformation of the country was presented by the Agency for Digital Italy.

The objectives of the three-year plan are based on the new European programming 2021-2027, on the principles of the eGovernment Action Plan 2016-2020 and on the actions foreseen by the eGovernment Declaration of Tallinn (2017-2021), whose indicators measure the level of digitization across the EU and detect the actual presence and use of digital services by citizens and businesses.

The **Three-Year Plan** for Public Administration Information Technology is focused on:

- The development of a digital society, where services put citizens and businesses at the center, through the digitization of the public administration which is the engine of development for the whole country;

- The promotion of sustainable, ethical and inclusive development, through innovation and digitization at the service of people, communities and territories, in compliance with environmental sustainability;

- The contribution to the diffusion of new digital technologies in the Italian productive fabric, encouraging standardization, innovation and experimentation in the field of public services.

The following are some examples of the objectives that must be pursued

- Improving the PA's ability to generate and deliver digital services through
- Increasing the level of adoption of the cloud enablement program, we have already implemented the migration of the thousands of data center sites to a hybrid cloud infrastructure for the Public Administrations and a limited number (<10) of data centers that fall within the scope of Poli Strategici Nazionali (- Strategic National Hubs).
- Sharing and re-use of data between PAs and re-use by citizens and businesses through
- Development of existing platforms, such as the digital payment platform (PagoPA), the digital identity (SPID), the National Registry (ANPR) and the electronic identity card (CIE) to improve the services offered to citizens and businesses by simplifying the administrative action. **Developers Italia** and **Designers Italia** platforms have been created to provide guidelines, examples and open source and design development kits for easy and rapid implementation - to enable developers, designers and technology providers to contribute to the evolution of digital public services, helping administrations deliver a modern user experience that is consistent and simple for all citizens;
- Increasing the level of supply and digitization of the Electronic Health Record with health documents by local health structures;
- Increasing in the number of Administrations served in **NoiPA** (Open data project for public administrations) and extension of the number of services offered by the platform (tax, social security, etc.) used. **NoiPA** has already developed technological solutions using opportunities provided by Blockchain in order to ensure data interoperability and provide the highest security of the managed information.
- Increasing the development of interoperability rules that are clearly defined and based on **API (Application Programming Interface**) to permit systems to communicate with each other;
- Increasing **Open Source** and **Open Government** as a method of collaborative development of digital public services, being more efficient and less costly. We have created dedicated tools – **DocsItalia** for the collection of documents related to public digital service and **ForumItalia** for citizens discussions on digital issues.
- Increasing the **reuse of data**, in particular on the territory, for sectors of interest, with the involvement (and support) of the regional and local government, activating dialogue with businesses and civil society, to better understand the meeting of data demand and supply as stimulator. Match this

objective with the actions envisaged in the context of the **Open Government Partnership** (OGP), also in line with the actions envisaged in this area;

- Increasing **Open data policies** with the spread of the culture of open data amongst administrations and businesses, based on inclusive actions aiming to foster an understanding of the potential linked with the reuse of data, at the same time avoiding any new or greater charges for the administrations

- Ensuring coordinated action by central, regional, and local administrations, for particular domains or ecosystems, aimed at opening data concerning the same object but containing different information, which complete each other (domain continuity)

- Promoting planned action to make some regional datasets available in all regions, so as to have coverage of all national territory (territorial continuity)

- Improving the **National Data Digital Platform (PDND)** as part of the activities aimed at optimising the national public information assets. The **National Data Digital Platform** develop and simplify the interoperability of public data between Public Administration through big data collection processing and sharing. It also optimises data analysis and knowledge generation processes. The idea is to have the world of the Public Administration up to the benefits offered by modern big data management and analysis platforms, acting along the following main lines:

1) significantly extend the value of the information assets of the Public Administration through the use of big data technologies that can create knowledge for those needing to make decisions and drastically reduce analysis time. The horizontal scalability of these technologies in fact makes it possible to extract information from the cross point of multiple databases and to process data in real time, making it possible to have several different analysis perspectives on a given phenomenon, in a timely manner;

2) foster and optimise the exchange of data between PA, minimising transaction costs for access and use. It will, in fact, be possible to overcome the structure of the agreements one-by-one, which will lead to multiple copies of the same data and allow for standardised access to data that is always up-to-date;

3) encourage the dissemination of open data and make its use more effective. Indeed, the **National Data Digital Platform** allows for the centralisation and redistribution of public data through the **Application Programming Interface**, guaranteeing the standardisation of formats and methods of re-use on data that is always up-to-date;

4) foster the exploratory analysis of data by teams of data scientists, both within the individual PAs and centrally, so as to improve knowledge of the

social phenomena. The analysis techniques used will also allow for the development of smart applications that exploit the regularity in the data to offer services to citizens, business and public administrations;

Effective use of big data would permit the public administration to measure the impact of policy making and public spending in the most varied of areas, including mobility, energy consumption, education, labor market, health and the fight against tax evasion and **corruption**.

The technological platform has reached an experimental stage, and, at present, can be used, amongst others, for the following study cases:

- National Registry monitoring dashboard;
- Statistical monitoring of access to qualified services for the Region of Tuscany;
- Monitoring of air quality for Rome capital city;
- Map of districts for Milan City Council

- Ensuring transparency in the creation and use of laws by **Lex Datafication Project.** The management of the immense patrimony of normative texts, be they primary laws or rules or secondary regulations, is fragmented and involves all phases: production, collection and management, publication for research and application.

In order to allow machine-to-machine sharing of legal texts, it is necessary to standardize the regulatory patrimony (international XML standard), along with introducing the use of machine learning and data science tools for the creation of modern search engines for information and content.

For example, this may lead to the production of services in the short term, such as the automatic generation of requests based on current European, national, and even local regulations.

Digital transformation of regulatory information assets should also involve the entire process drafting a law - the introduction and adoption of collaborative digital tools and processes (including those typical of the open-source world) to allow the drafting process not only to be digital by default but to take place openly and transparently, involving citizens and businesses. Such tools would facilitate the sharing of legislative proposals in public consultation and the structured collection of contributions and proposals for amendments by citizens and businesses.

We have also developed **io.italia.it**, a highly innovative application that will allow citizens to easier interact with public services such as to receive messages, documents and deadline notifications, request information and certificates, set their preferences and pay the public administration in a simple and intuitive way.

As mentioned before, I will also share with you the **project implemented by the Scuola Superiore di Sant'Anna di Pisa.**

**The Predictive Justice** is an ambitious long-term project built on an innovative approach and philosophy.

The approach is based on the assumption that only the combination of different expertise and tools in coherent and integrated pipelines can deliver effective results quickly. For this reason, the teams of each 'sub-project' are interdisciplinary and the individual sub-projects are also scientifically and operationally autonomous projects. It is precisely the idea that the advancement of knowledge with its application implications offers building blocks that can be organised modularly in a variety of ways that has determined the organization of the work into building-block projects.

The ambitions are multiple and transversal: from the attempt to "export" knowledge, techniques, and solutions across disciplines (e.g. from omics to legal data mining), to the coupling of protocols and software to automate the pseudonymisation of texts, or to the creation of innovative tools for querying legal materials through their automatic annotation, to the construction of predictive tools based on data science and Artificial Intelligence, to the attempt to offer comprehensible explanations on the functioning of the tools used and adapt them to the various end-users' needs/abilities.

All these steps are obviously articulated in full coherence with the corresponding regulatory and ethical frameworks, in the belief that regulatory and ethical profiles in research are central and go beyond mere adherence to norms.

The objective of DATABASE Legal Document Management System (LDMS) is to develop computerised tools for the analysis of case law and legal texts in order to:

➢ Allow immediate access to automatically annotated and semantically enriched case law precedents, creating innovative case law archives capable of identifying the types of sentences within the decision and placing them in the context of reference in order to favour the analysis and creation of the correct legal arguments and to identify the determining factors in decisions

➢ Create user interfaces suitable for consultation by the various professionals involved and stakeholders

➢ Generalise the results obtained ideally on any legal topic.

To achieve the above objectives, Legal Document Management System (LDMS) is divided into the following tasks:

- Exploratory Data Analysis (EDA)
- Extract, Transform, Load (ETL)
- Data Lake
- Data Labeling UI
- Legal Semantic Search Engine

Exploratory Data Analysis (EDA) is used to analyze and investigate data sets and summarize their key characteristics. It uses data visualization methods to discover patterns, spot anomalies (missing values, outliers, duplications), test a hypothesis, or check assumptions. It provides a better understanding of the variables in the datasets analyzing national case law distributions/trends and offers an effective tool for legal comparisons.

Extract, Transform, Load (ETL) - it automatically converts PDF files and legal document metadata into the flexible LS-JSON data format. LS-json can capture all aspects of a legal document and is used to add entity labels and relationships on different portions of text.

In malformed text, machine learning models have difficulty capturing context. The text pre-processing phase of the ETL module will apply a variety of techniques to convert raw text data into clean data and standardized sequences, which can improve the performance of predictive methods. The results of the predictive models trained on clean data will be compared to those obtained from the raw data in order to assess the impact of the cleaning process on model quality.

The Data Lake is based on a flexible data format and schema-free NoSql storage with data versioning and ML reproducibility. It can handle large datasets and track ML experiments, model parameters, metrics and artifacts. Thanks to the integrated tracking system, the Data Lake guarantees the scientific integrity of the experiments that can be reproduced for validation and future improvements of the models.

Data Labeling is aimed at supervising learning methods and assumes that the annotations provided for model training are correct. It introduces a hierarchical scheme of annotators and a user interface tool that guides experts through the annotation process to obtain high quality dataset. The team of domain exporters who will label the judgments to be used for algorithm training is composed of

law students led by specialized attorneys in the different subjects of jurisprudence who will evaluate the qualities of the annotation based on taking into account the inter-label correlation and their professional experience. An efficient text annotation tool will be used to quickly annotate and normalize entities meeting quality requirements.

Legal Semantic Search Engine is used to date, legal databases are queried by users using keywords. The innovation that Legal Semantic Search Engine is aiming at is radical: transforming common case law collections into 1) automatically semantically annotated databases and 2) queryable not only by keywords but also by types of phrases and semantic similarity 3) in the context of the individual decision and 4) providing a "value" on the quality of the detected phrase. Such results are not only an innovative aid for all legal practitioners but also for policy makers. Little effort has been devoted to exploiting case law data for regulatory policy development to inform potential reforms. Legal Semantic Search Engine is important because it develops these needed tools for research in legal documents to better inform policy makers as well.

***
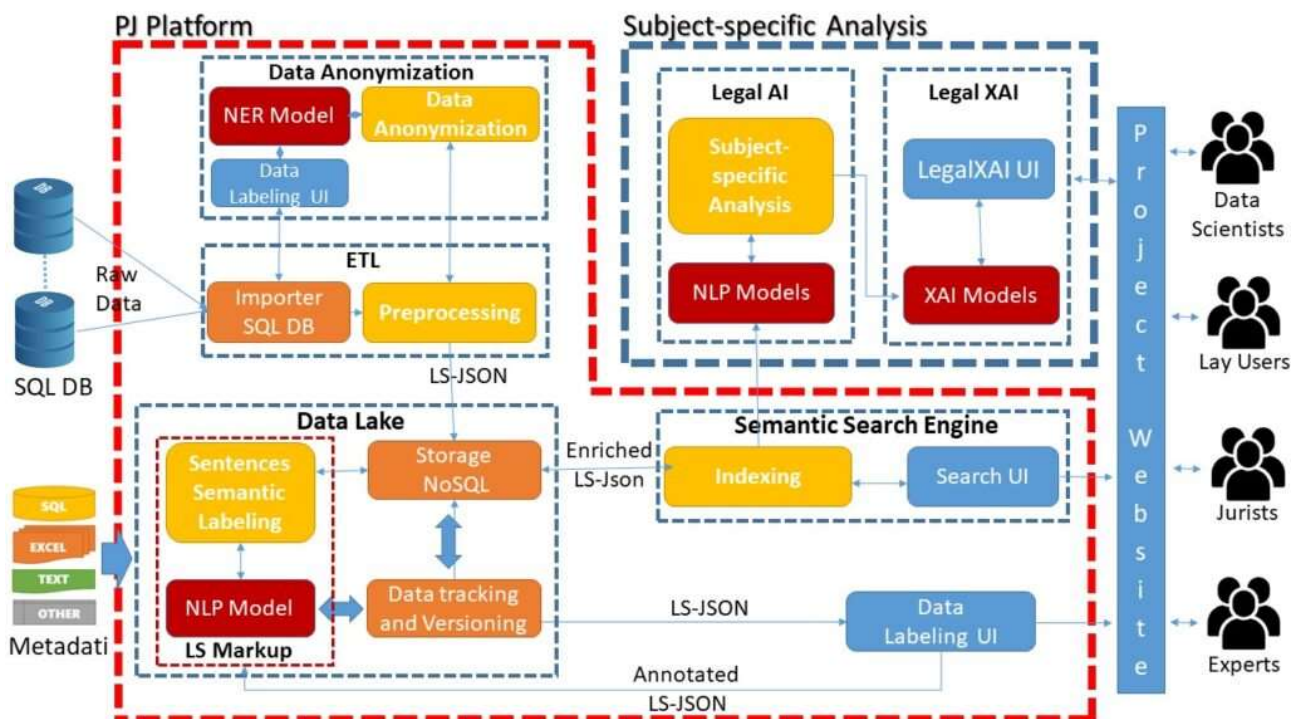
The expression Predictive Justice is often associated with the idea that it is possible to replace the decision-maker with automatic systems, and today with Artificial Intelligence based tools. Although we are convinced that some rare instances, being simple and repetitive, can benefit from certain levels of automation, we remain convinced that human decision making rather than being replaced should be assisted by technology.

So, are emerging technologies helping win the fight against corruption?
It appears that Information and Communication Technology can:
- ensure more efficiency, transparency and accountability in the public sector;
- facilitate advocacy and citizen participation as well as closer interaction between government and citizens;
- increase integrity, simplify procedures and over time reduce or undercut opportunities for bribery; and
- genuinely **support anti-corruption** action by impacting on public scrutiny in numerous ways.

Thank you very much for your kind attention!