



Federal Court of Accounts - Brazil (TCU)

# INITIATIVES IN DATA SCIENCE

PAS-meeting, Budapeste, April 2018  
Dagomar Lima





# PRESENTATION TOPICS

1. Why did TCU see increased efforts in Data Science as important?  
How did TCU get started to use Data Science?
2. Examples

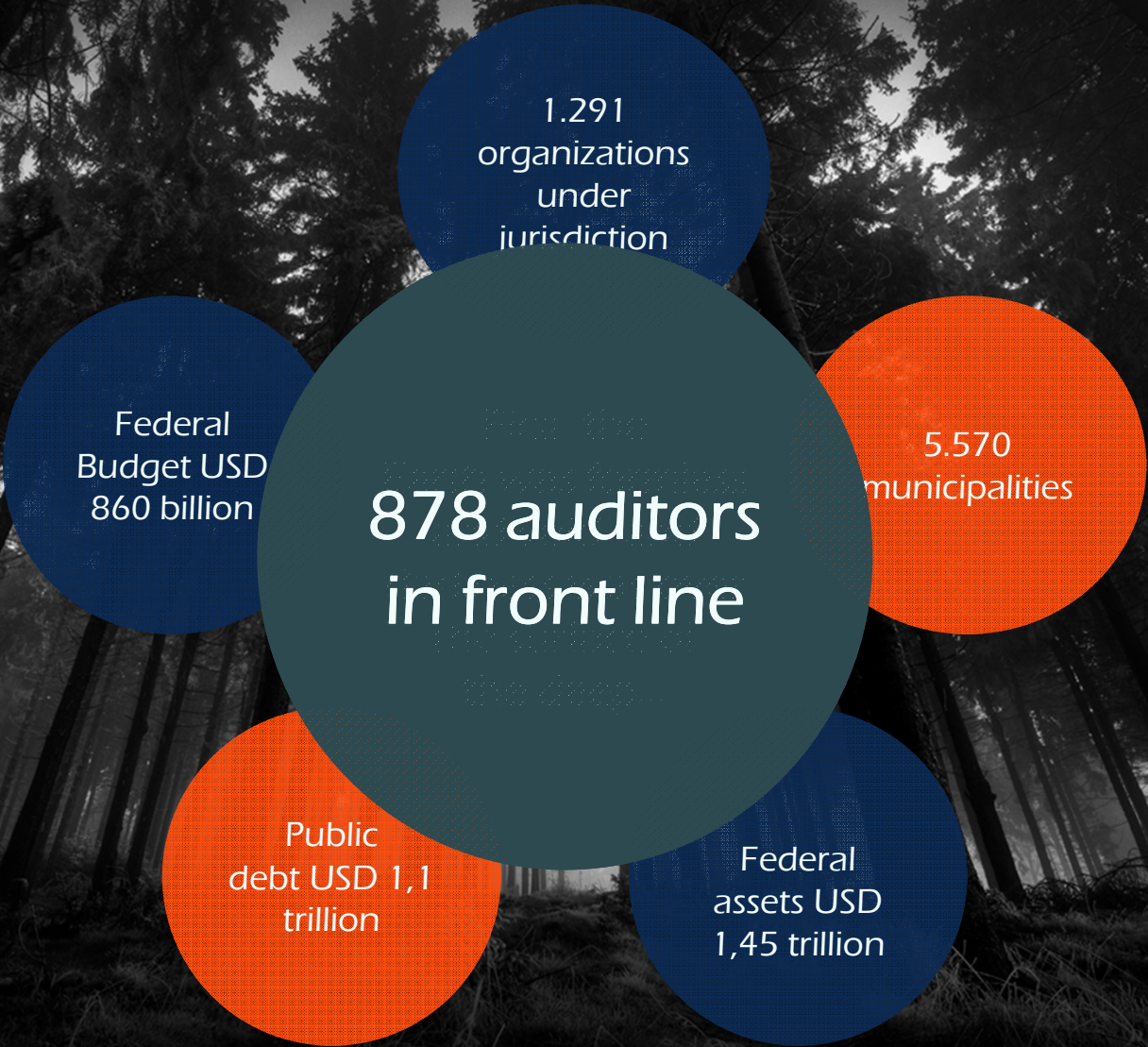


## **Why did TCU see increased efforts in Data Science as important?**



**To organize the information from Public Administration and make it accessible and useful to auditors**





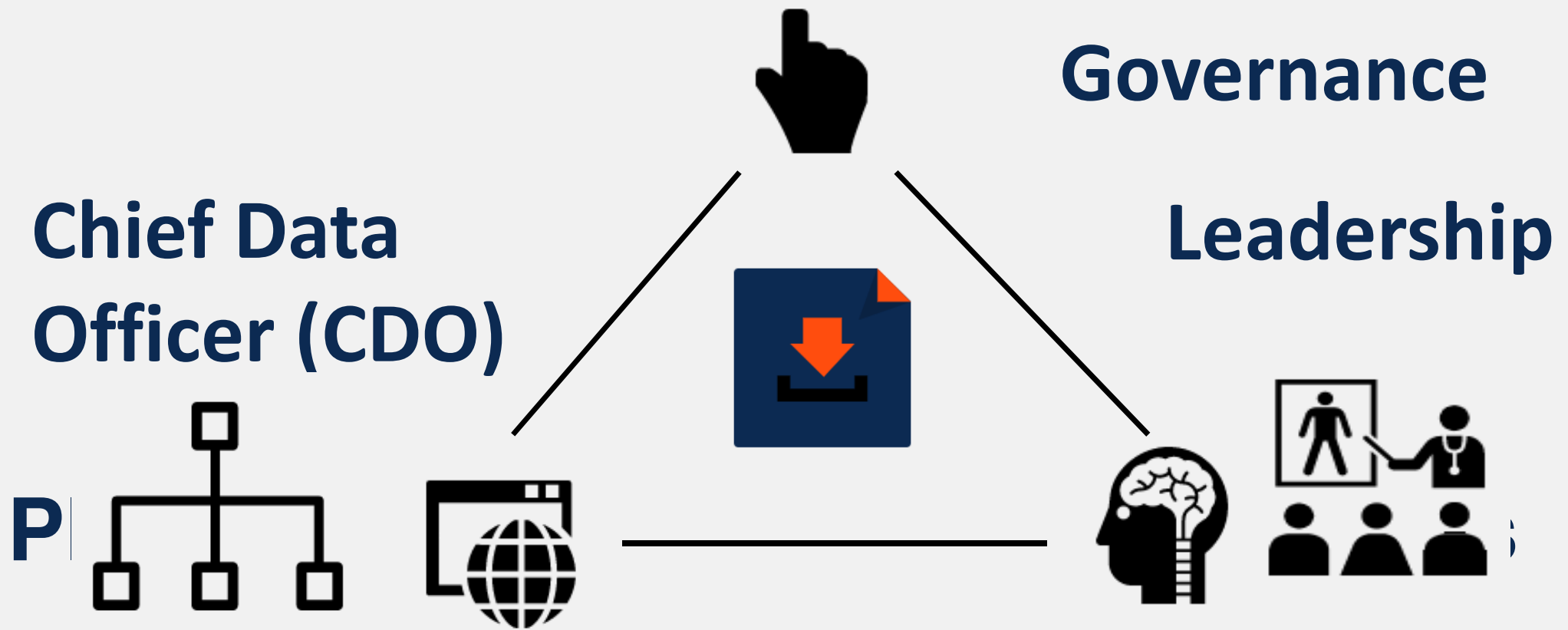




Data analysis  
is for  
everyone



# TCU Strategy





# Governance

# 2017

## Business Unit

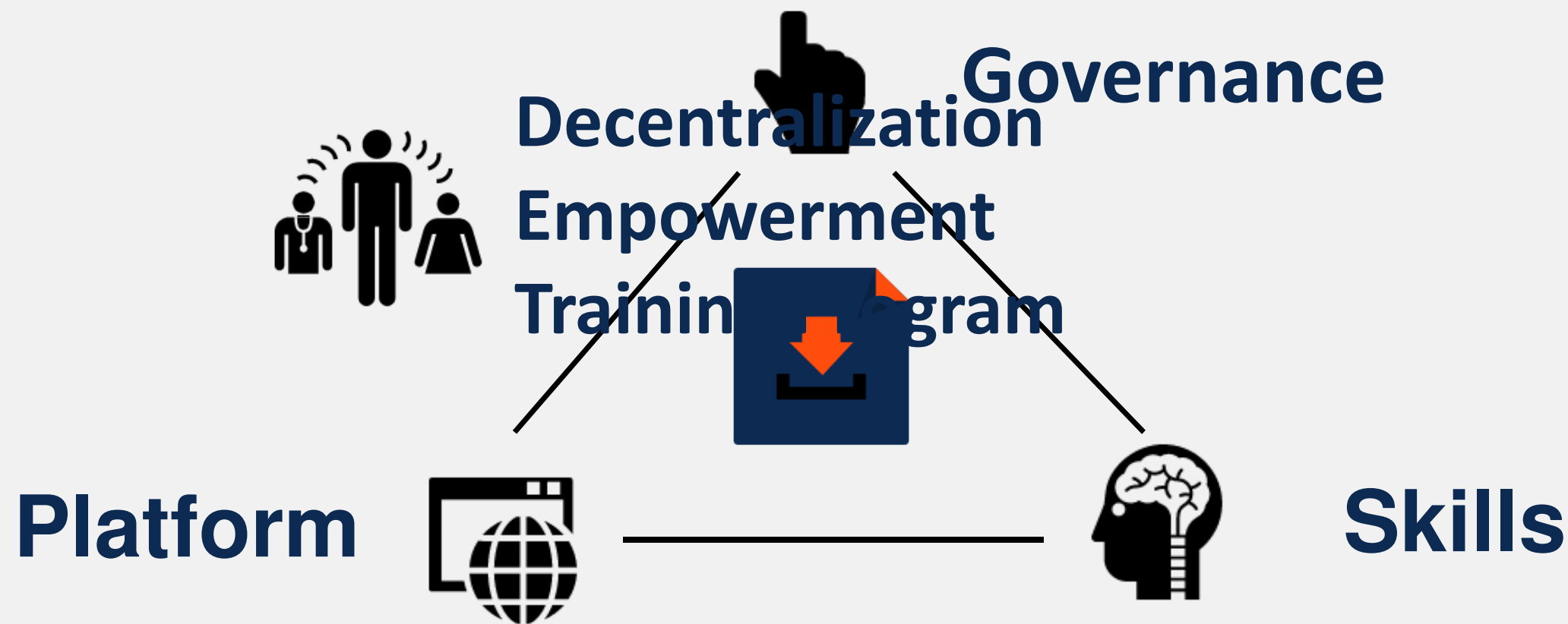
# Audit Information Management - Data Analysis

# CDO

# Systems Management



# TCU Strategy

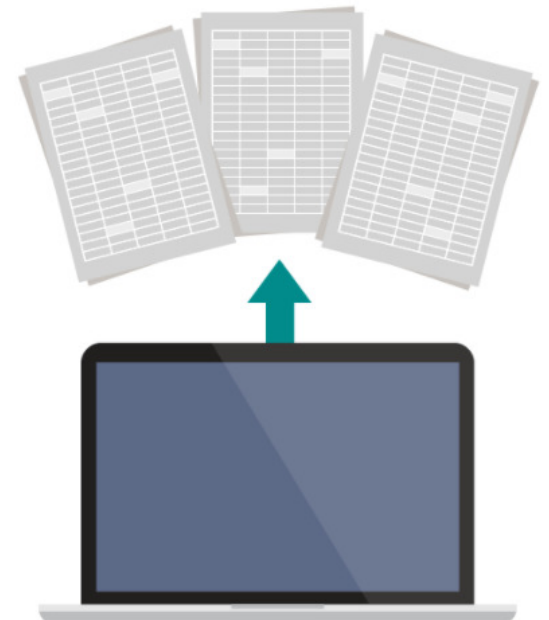
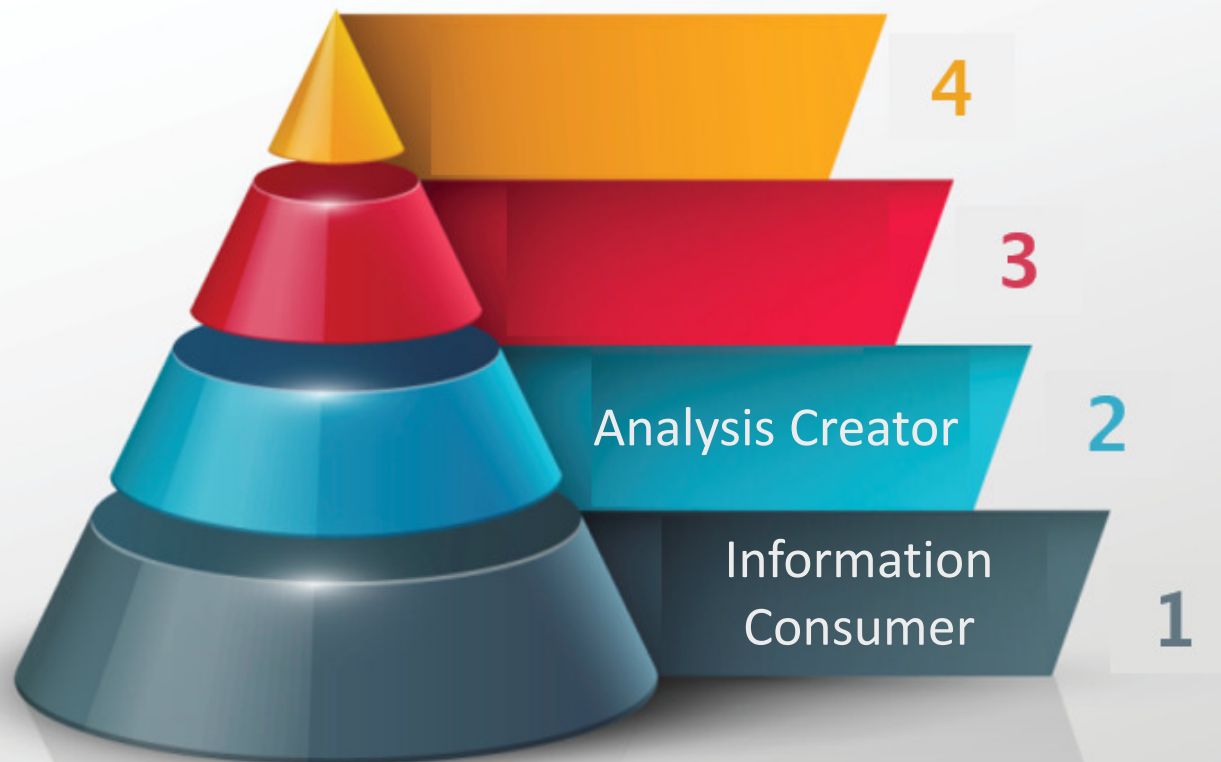




# Data analysis – four levels

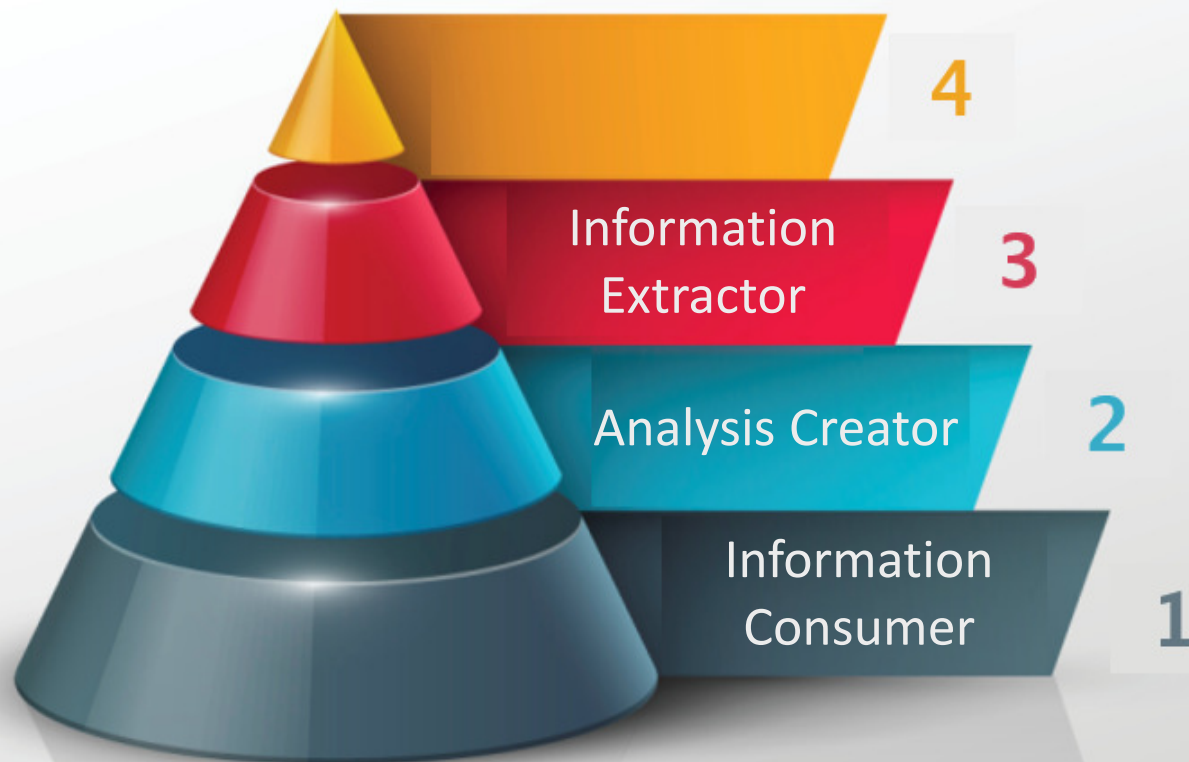


# Data analysis – four levels





# Data analysis – four levels

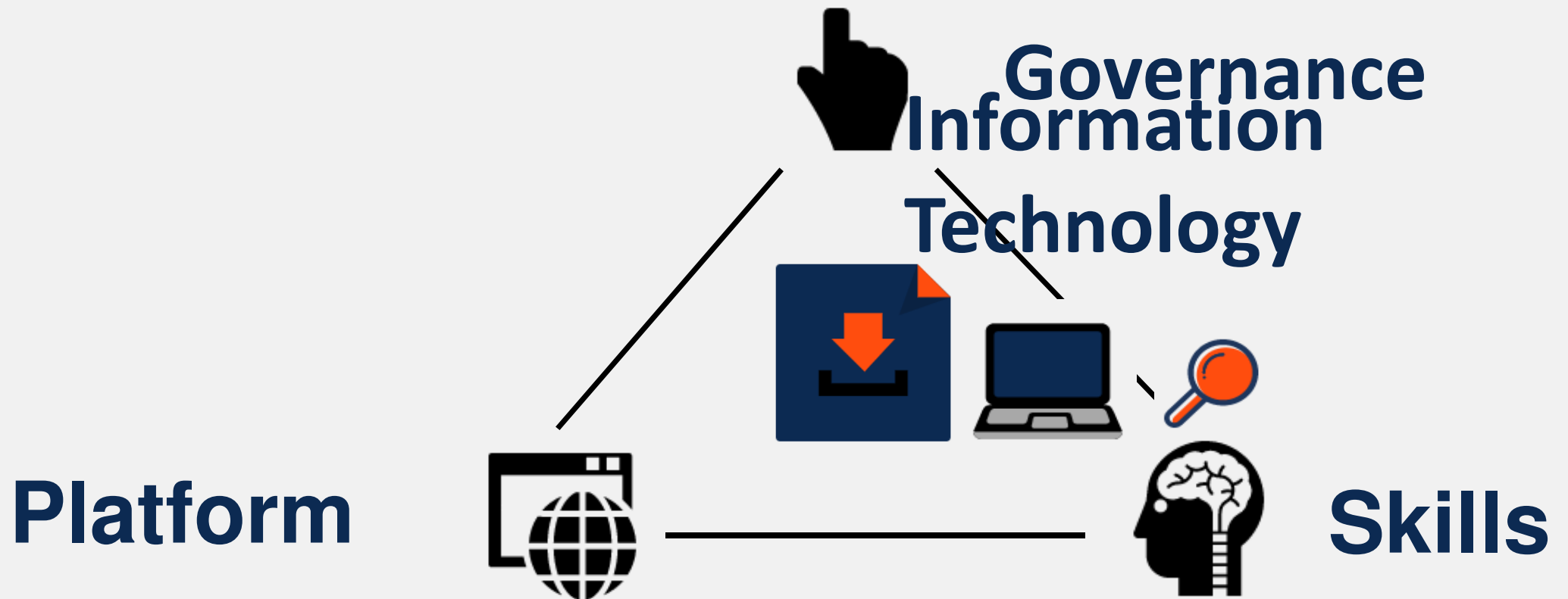


# Data analysis – four levels

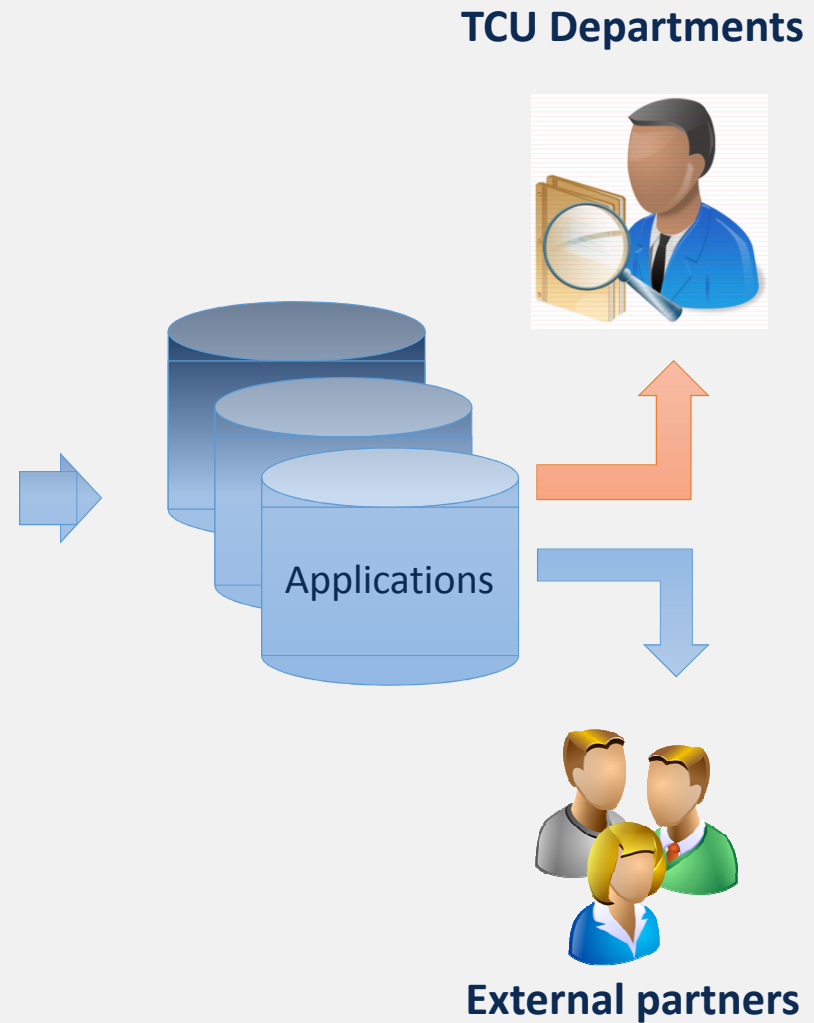
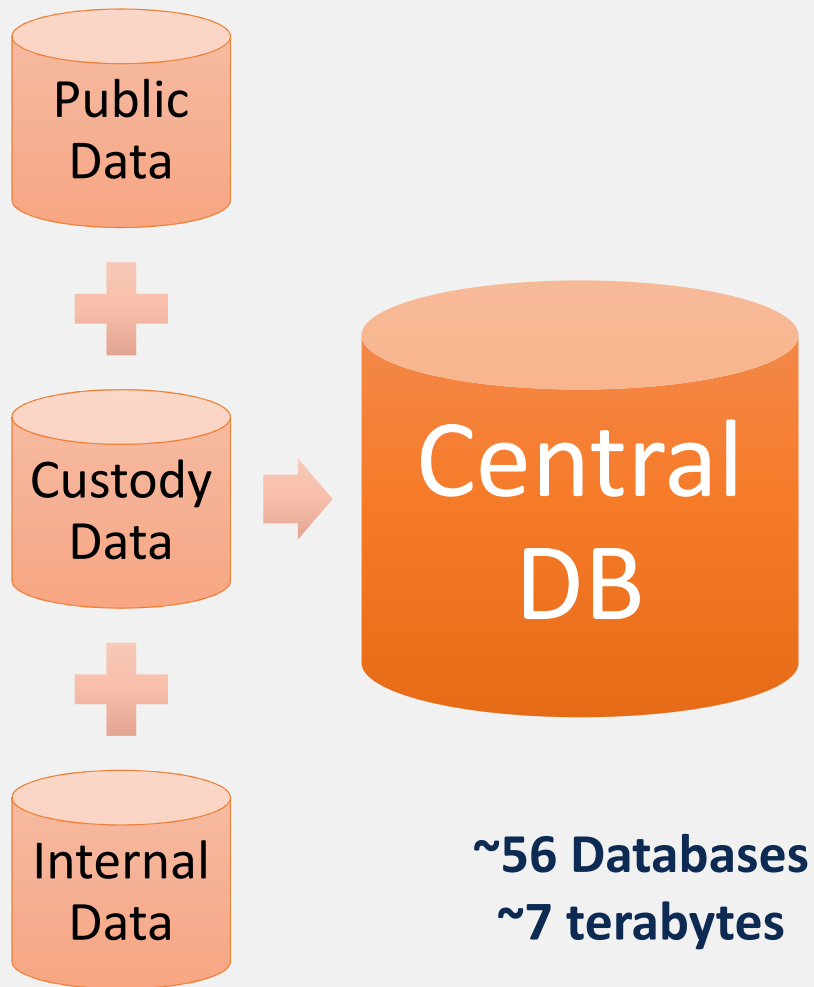




# TCU Strategy



# Platform







## 2

# Examples



**Investigation - Corruption in Petrobras**

**Compliance - Robot Alice**

**Boundry – Predictive analysis**

**Performance Auditing - Program of Access to Technical Education and Employment (Pronatec)**



**Corruption  
in Petrobras  
("Lava-Jato")**

**156 contracts (Years 2002-2015 / Values over  
US\$ 31 million)**

**40 contracts with participation of companies  
that formed cartel (police information)**

**Eight econometric models with 16 variables**

**Estimate of the damage suffered by Petrobras, related to the reduction of the discount in the bids due to the action of a cartel (overcharge)**





Corruption  
in Petrobras  
("Lava-Jato")

## Conclusions

**17% overcharge** whenever a firm from the the cartel won a bid.

Damage around **US\$ 3 billion**

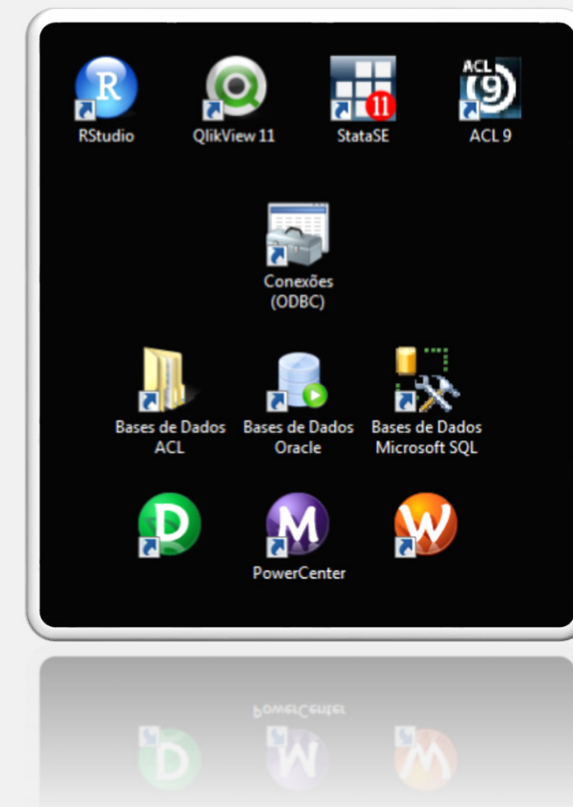
The probability that the cartel does not influence the bid discount **is less than 1%**

Estimate of the damage suffered by Petrobras, related to the reduction of the discount in the bids due to the action of a cartel (overcharge)



## Control Information Lab (LabContas)

- Virtual machine with data and analytics tools
- Around 600 internal users and a 50 external users
- Partners – 17 Courts of Accounts and other control institutions



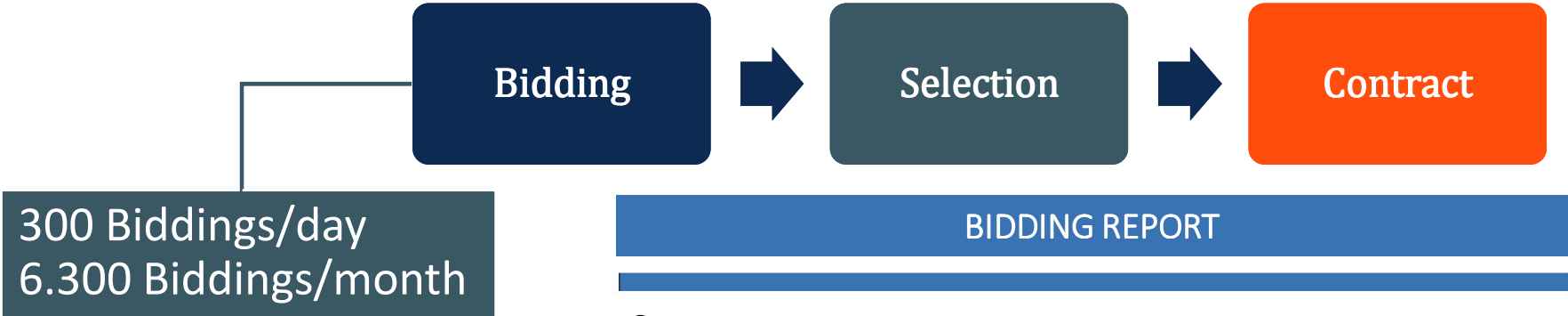
# Robots Alice and Sofia



LabContas



# Alice



### Summary

On 10/7/2016, were published 386 biddings. 30 of them are from organizations of your interest

Licitações							
UASG	Unidade Responsável pela Licitação	UF	Modalidade	Edital	Estimativa da Unidade*	Materialidade	Alertas
<a href="#">120625</a>	COMANDO DA AERONAUTICA	DF	Pregão	<a href="#">35/2016</a>	R\$ 8.760.470,14	\$ \$ \$ \$ \$ \$	
<a href="#">590001</a>	MINISTERIO PUBLICO DA UNIAO	DF	Pregão	<a href="#">30/2016</a>	R\$ 49.975,71	\$ \$	
<a href="#">30001</a>	TRIBUNAL DE CONTAS DA UNIAO	DF	Pregão	<a href="#">81/2016</a>	R\$ 47.436,98	\$	
<a href="#">200009</a>	MINISTERIO PUBLICO DA UNIAO	DF	Pregão	<a href="#">68/2016</a>	Não Identificado		
<a href="#">160067</a>	COMANDO DO EXERCITO	DF	Pregão	<a href="#">2/2016</a>	R\$ 366.855.031,64	\$ \$ \$ \$ \$ \$	
<a href="#">395001</a>	EMPRESA DE PLANEJAMENTO E LOGÍST...	DF	Pregão	<a href="#">2/2016</a>	R\$ 928.590,40	\$ \$ \$ \$ \$ \$	
<a href="#">120006</a>	COMANDO DA AERONAUTICA	DF	Pregão	<a href="#">45/2016</a>	R\$ 2.960.970,00	\$ \$ \$ \$ \$ \$	
<a href="#">110404</a>	MINISTERIO DEFESA	DF	Pregão	<a href="#">26/2016</a>	R\$ 1.677.288,47	\$ \$ \$ \$ \$ \$	
<a href="#">393001</a>	AGENCIA NACIONAL DE TRANSPORTES ...	DF	Pregão	<a href="#">22/2016</a>	R\$ 1.045.901,10	\$ \$ \$ \$ \$ \$	
<a href="#">10001</a>	CAMARA DOS DEPUTADOS	DF	Pregão	<a href="#">122/2016</a>	R\$ 498.243,71	\$ \$ \$ \$ \$ \$	
<a href="#">410003</a>	MINISTERIO DAS COMUNICACOES	DF	Pregão	<a href="#">9/2016</a>	R\$ 1.432.543,37	\$ \$ \$ \$ \$	
<a href="#">380918</a>	FUNDO DE AMPARO AO TRABALHADOR	DF	Pregão	<a href="#">22/2016</a>	R\$ 474.038,40	\$ \$ \$ \$ \$	
<a href="#">110001</a>	PRESIDENCIA DA REPUBLICA	DF	Pregão	<a href="#">28/2016</a>	R\$ 372.600,00	\$ \$ \$ \$ \$	
<a href="#">10001</a>	CAMARA DOS DEPUTADOS	DF	Pregão	<a href="#">124/2016</a>	R\$ 208.830,00	\$ \$ \$ \$ \$	
<a href="#">50001</a>	SUPERIOR TRIBUNAL DE JUSTICA	DF	Pregão	<a href="#">123/2016</a>	R\$ 105.153,84	\$ \$ \$ \$ \$	



## Análise Geral

### Filtros aplicados

Nenhum

« Limpar filtros »

Secretaria do TCU > Órgão/Entidade > Unidade

- ☐ Estadual
- ☐ Federal
- ☐ Federal-DF
- ☐ Municipal
- ☐ Secex-AC
- ☐ Secex-AL

### Esfera

☐ Municipal ☐ Estadual ☐ Federal

### UF > Município do Licitante

- ☐ AC ☐ DF ☐ MT ☐ RJ ☐ SE
- ☐ AL ☐ ES ☐ PA ☐ RN ☐ SP
- ☐ AM ☐ GO ☐ PB ☐ RO ☐ TO
- ☐ AP ☐ MA ☐ PE ☐ RR
- ☐ BA ☐ MG ☐ PI ☐ RS
- ☐ CE ☐ MS ☐ PR ☐ SC

### Poder

☐ Executivo ☐ Judiciário ☐ Legislativo

### Órg Sup > Órg Vinc > Órg > Unid

- ☐ ADVOCACIA-GERAL DA UNIAO
- ☐ AGENCIA DE DESENVOLVIMENTO DA AM
- ☐ AGENCIA DE DESENVOLVIMENTO DO NOI
- ☐ AGENCIA ESPACIAL BRASILEIRA - AEB
- ☐ AGENCIA NAC. DE TRANSPORTES AQUAVI
- ☐ AGENCIA NACIONAL DE AGUAS
- ☐ AGENCIA NACIONAL DE AVIACAO CIVIL

### Ano da Publicação do Edital

☐ 2... ☐ 2... ☐ 2... ☐ 2... ☐ 2... ☐ 2...

### Visão por Tipologia

### Georreferenciamento

### Detalhes dos Editais

### Detalhes das Tipologias

### Detalhes dos Fornecedores



12.655

Total de editais em risco na base

12.655 Editais em risco selecionados



29.030.272.652,73

Valor dos editais em risco na base (R\$):

29.030.272.652,73 Valor dos editais em risco selecionados (R\$):



1.284.204.949.811,14

Valor dos editais na base (R\$):

1.284.204.949.811,14 Valor dos editais selecionados (R\$):

Q Todos os campos

Q UASG

Q Nº Edital Ano

Q Tipologia

Q Objeto do Edital

Q CPF/CNPJ

Tipologia	Ocorrências	Editais	Valor Estimado do Risco (R\$)
Licitantes com contadores em comum	5.635	3.324	4.910.440.094,22
Licitantes com telefone/e-mail em comum	3.492	2.008	3.727.008.013,13
Exigência de carta de credenciamento emitida pelo fabricante	2.858	1.673	4.732.878.692,13
Licitante Único	2.841	2.841	2.486.174.037,71
Licitantes com Sócios com Parentesco	2.670	1.273	1.910.155.233,43
Exigência de comprovação de quadro permanente sem permitir contrato de prestação de serviços	2.077	1.671	6.287.385.712,76
Licitantes cujo sócio é parente de sócio de empresa proibida de contratar	1.332	584	233.931.435,74
Licitantes com ex-Sócios em Comum	975	718	1.626.090.041,32
Exigência de capital social ou patrimônio líquido integralizados	742	698	3.721.311.178,76
Licitantes com Sócios em Comum	721	369	960.715.296,06

### Valor Estimado do Risco (R\$)

Órgão Superior



### Ocorrências por Modalidade





Predictive  
analysis

Can we  
**PREDICT** or  
anticipate our  
**FUTURE** by  
learning  
**FROM** our  
**PAST?**

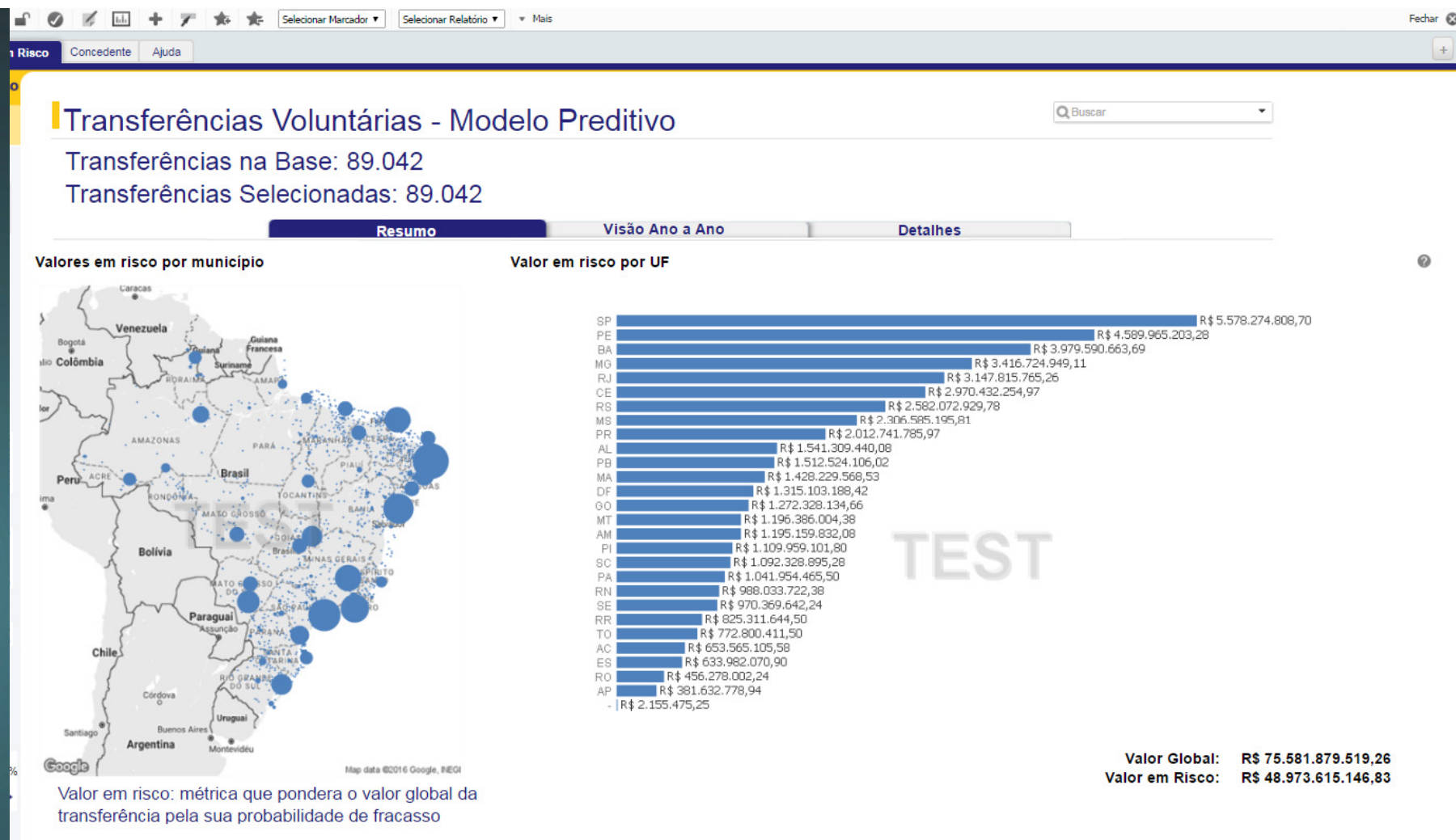
PAST

FUTURE

Through **PREDICTIVE ANALYSIS** it is possible to issue **ALERTS**, based on risk analysis, both to **SUPPORT** preventive **ACTIONS** and to **SELECT SUBJECT MATTERS** for audits and other engagements



Predictive  
audits

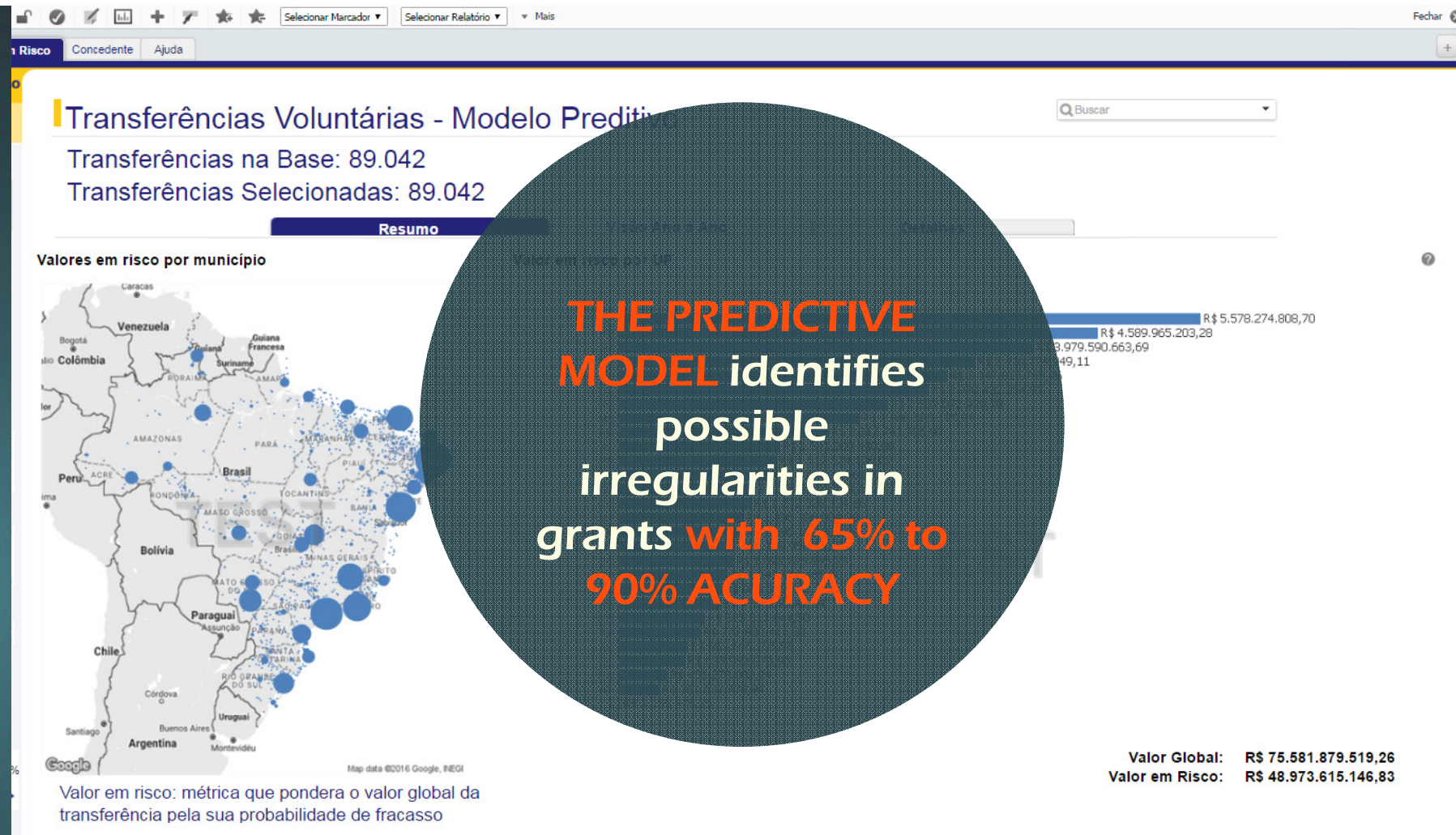


The **RISK ANALYSIS** of grants to subnational entities allowed the elaboration of a catalog of typologies of irregularities





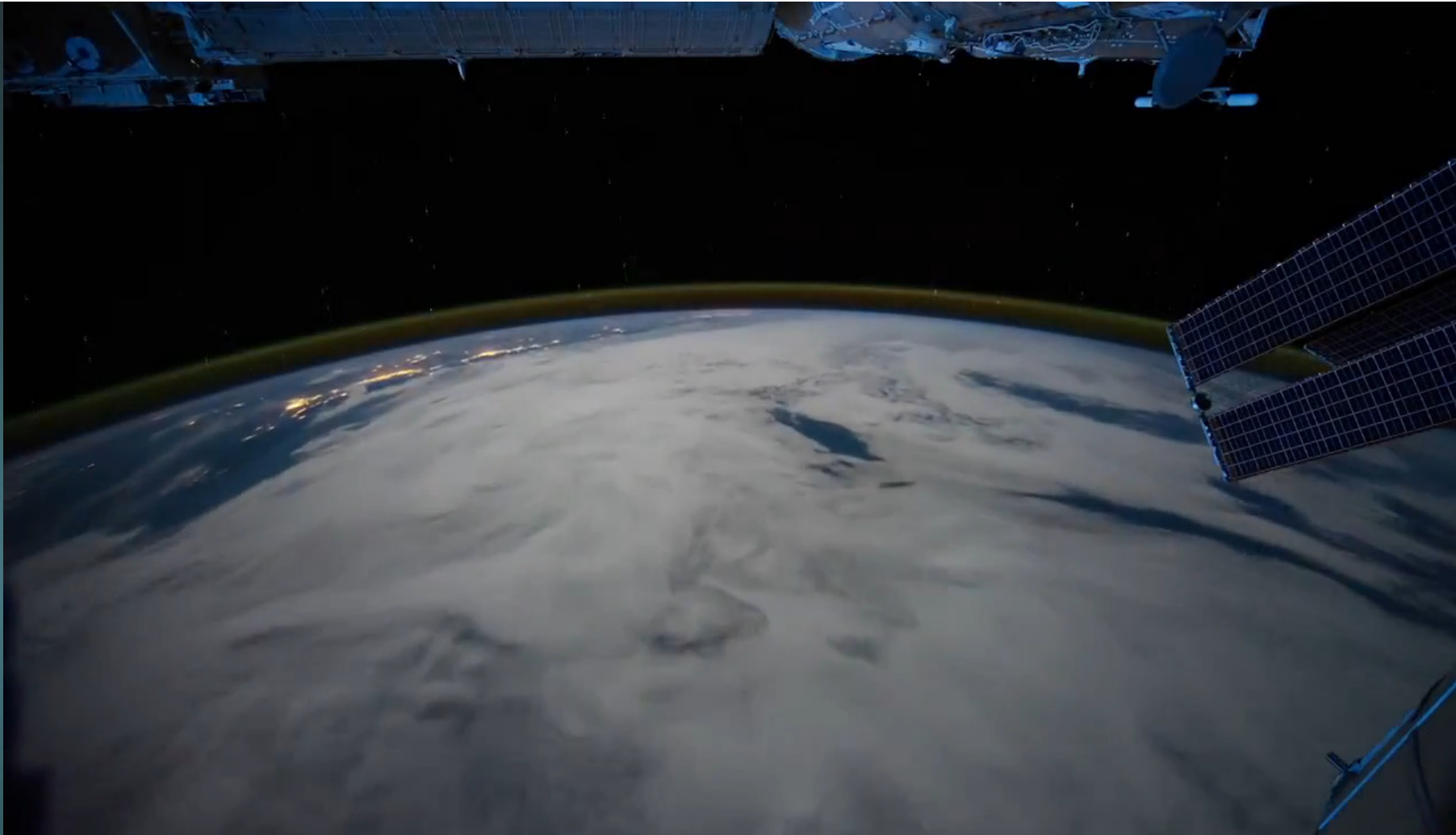
Predictive  
audits



CRISP-DM, C50, R, 24 variables, 10 thousand cases. Failure=do not deliver public policy  
Audits in 9 states, 51 agreements.  
Limitations



**Geocontrol**



Set of technologies used for **COLLECTING, PROCESSING, ANALYSING and PRESENTANTING GEORREFERENCED INFORMATION** such as: remote sensing | multicriteria decision analysis of geographic information | geographic information systems (GIS)





Geocontrol



Geoprocessing, through Spatial MCDA, **HELPS** public administration to **DETERMINE BEST ROUTE** (linear works) and **OPTIMIZED LOCATION** (point works)





Geocontrol

**CHANGE  
DETECTION**  
between photos  
from the **SAME**  
**LOCATION** at  
**DIFFERENT**  
**TIMES**



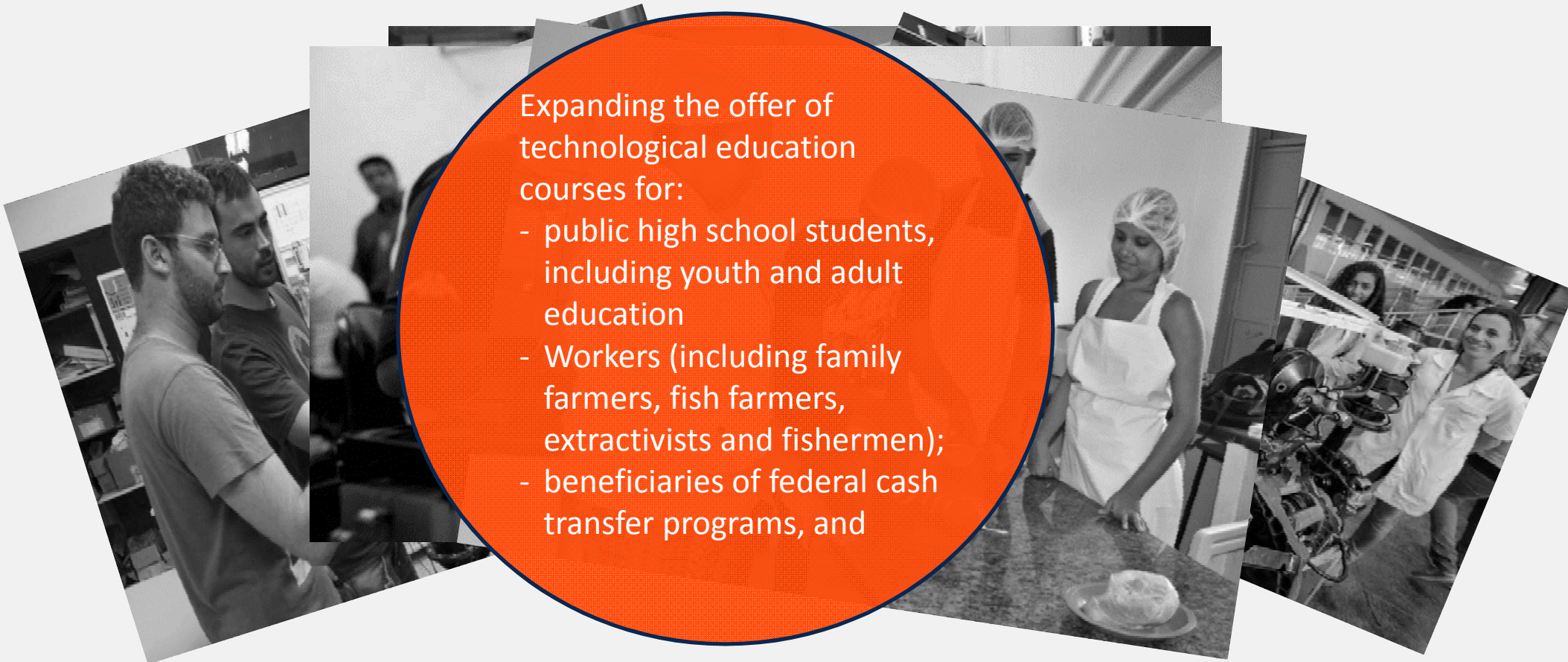
Audit the construction of the North-South Railway with the support of aerial image collected by **Unmanned Aerial Vehicle** (UAV) and overlap with the engineering project images



# Performance Auditing - Pronatec

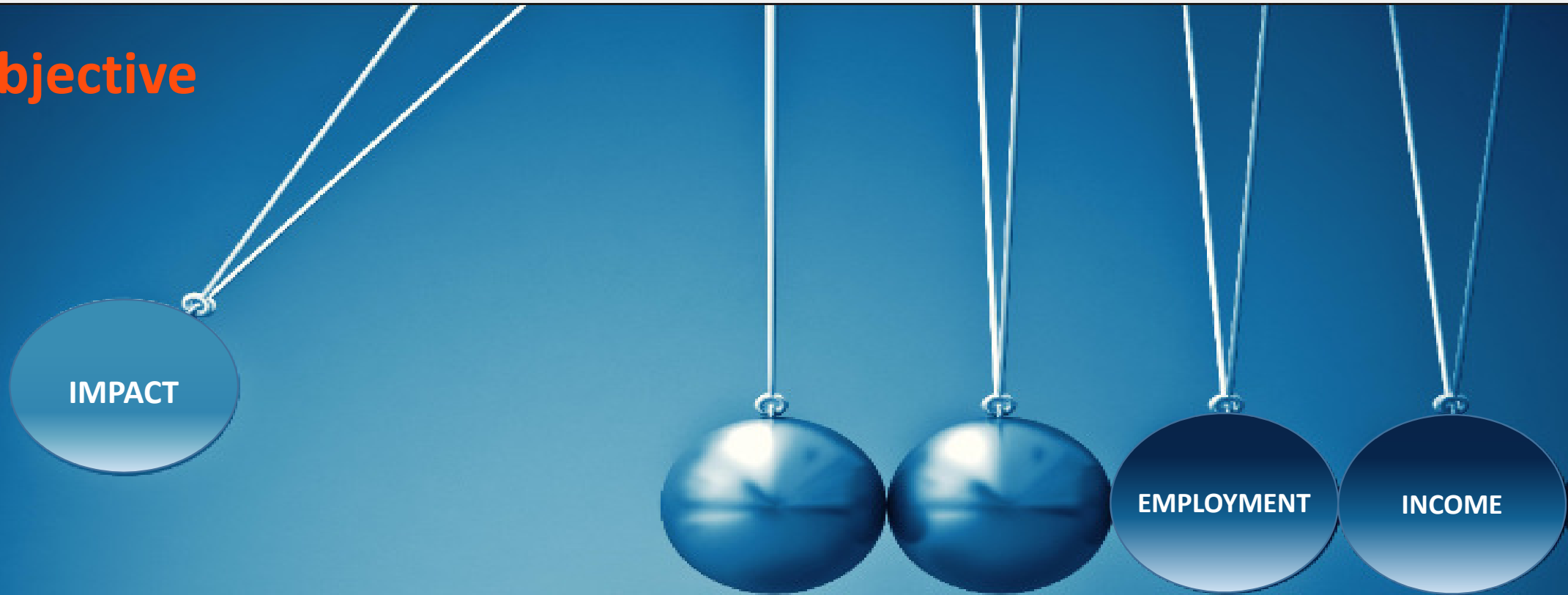
Expanding the offer of technological education courses for:

- public high school students, including youth and adult education
- Workers (including family farmers, fish farmers, extractivists and fishermen);
- beneficiaries of federal cash transfer programs, and



# Performance Auditing - Pronatec

Objective



# Performance Auditing - Pronatec

## Method

Treatment group



Control group





# Performance Auditing - Pronatec

## Method

### Sample - 2012

Scholarship-2012		
Group	Total	Percentage (%)
Control	16.179	26.31
Treatment	29.345	73.69
Total	45.524	100

### Econometric Model

$$DD = \{E[Y_{it}|P_i = 1, T_t = 1] - E[Y_{it}|P_i = 1, T_t = 0]\} - \{E[Y_{it}|P_i = 0, T_t = 1] - E[Y_{it}|P_i = 0, T_t = 0]\}$$

# Performance Auditing - Pronatec

## Method

$$Y_{it} = \beta_0 + \beta_1 P_i \times T_t + \beta_2 P_i + \beta_3 T_t + \varepsilon_{it}$$

What means :

$Y_{it}$  represents the probability of formal job or the wage of individual “i” in the period “t”;

$P_i$  is the treatment dummy variable, equal to “1” if the person participates in Pronatec and “0” if the person do not participate;

$T_t$  is a year dummy, equal to “0” in 2011 and “1” in 2013; and

$\varepsilon_{it}$  is the random error term not correlated with the regressors.

# Performance Auditing - Pronatec

## Method Observable characteristics used in the calculation of the Propensity Score

Control Variables	Description	Source	Control Variables	Description	Source
Average_wage	Average wage of the months that the person worked	RAIS 2011	group n	Dummies that represent the 90 groups of job types	RAIS 2011
Gender	Dummy equal to "1" if it is male, and "0" if it is female	Pronatec 2012	uf_n	Dummies that represent the 27 Brazilian States	RAIS 2011
Age	Age in years	Pronatec 2012	population	Total population of the city	IBGE 2011
Age2	Age squared	Pronatec 2012	working_age	Population percentage in working age (25 to 65 years)	DATASUS 2011
High_school	Dummy equal to "1" if the person has completed High School, and 0, otherwise	Pronatec 2012	elderly	Elderly population percentage (over 65 years)	DATASUS 2011
Higher_education	Dummy equal to "1" if the person has completed University, and "0" otherwise	Pronatec 2012	young	Young population percentage (below 25 years)	DATASUS 2011
experience	Experience in years	RAIS 2011	Political variables	Political variables of the city	TSE 2008
experience2	Experience squared	RAIS 2011			
tempultemp	Time in the last job	RAIS 2011			

# Performance Auditing - Pronatec

## Results

Impact on the employability\* of beneficiaries of PRONATEC courses

	2011 -2013	2011-2014	2011-2015
General treatment group (Everyone)	Not conclusive	Not conclusive	2.8%
MDS-restricted treatment group	6.5%	4.9%	2.7%
Non-MDS - restricted treatment group	- 2.3%	- 3.8%	3.5%

\* There wasn't statistically significant impact on income



# Strategy Innovation

Thank you  
**for your**  
attention

Vision  
Creativity

Support

Solution

BRAZILIAN FEDERAL COURT OF ACCOUNTS

Wesley Vaz Silva – wesleyvs@tcu.gov.br

Dagomar Lima – dagomarhl@tcu.gov.br