



Quality Control Practices at Central Bank of Brazil

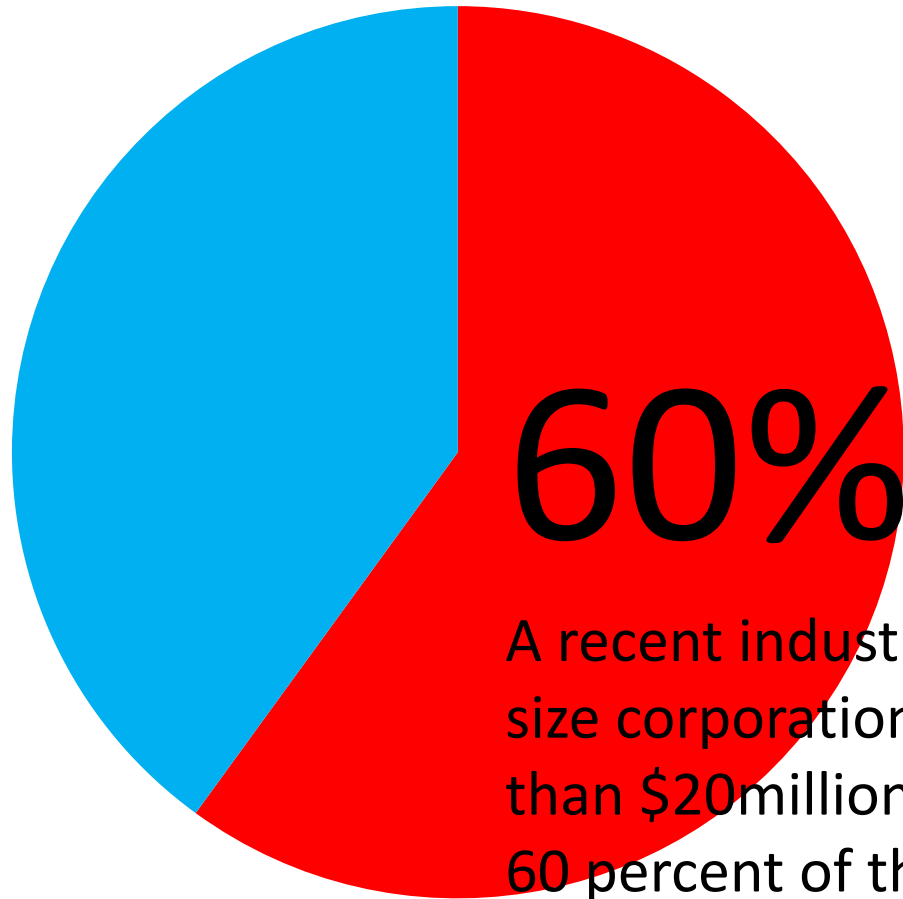
Paula Oliveira

Data Quality – an issue for everyone!



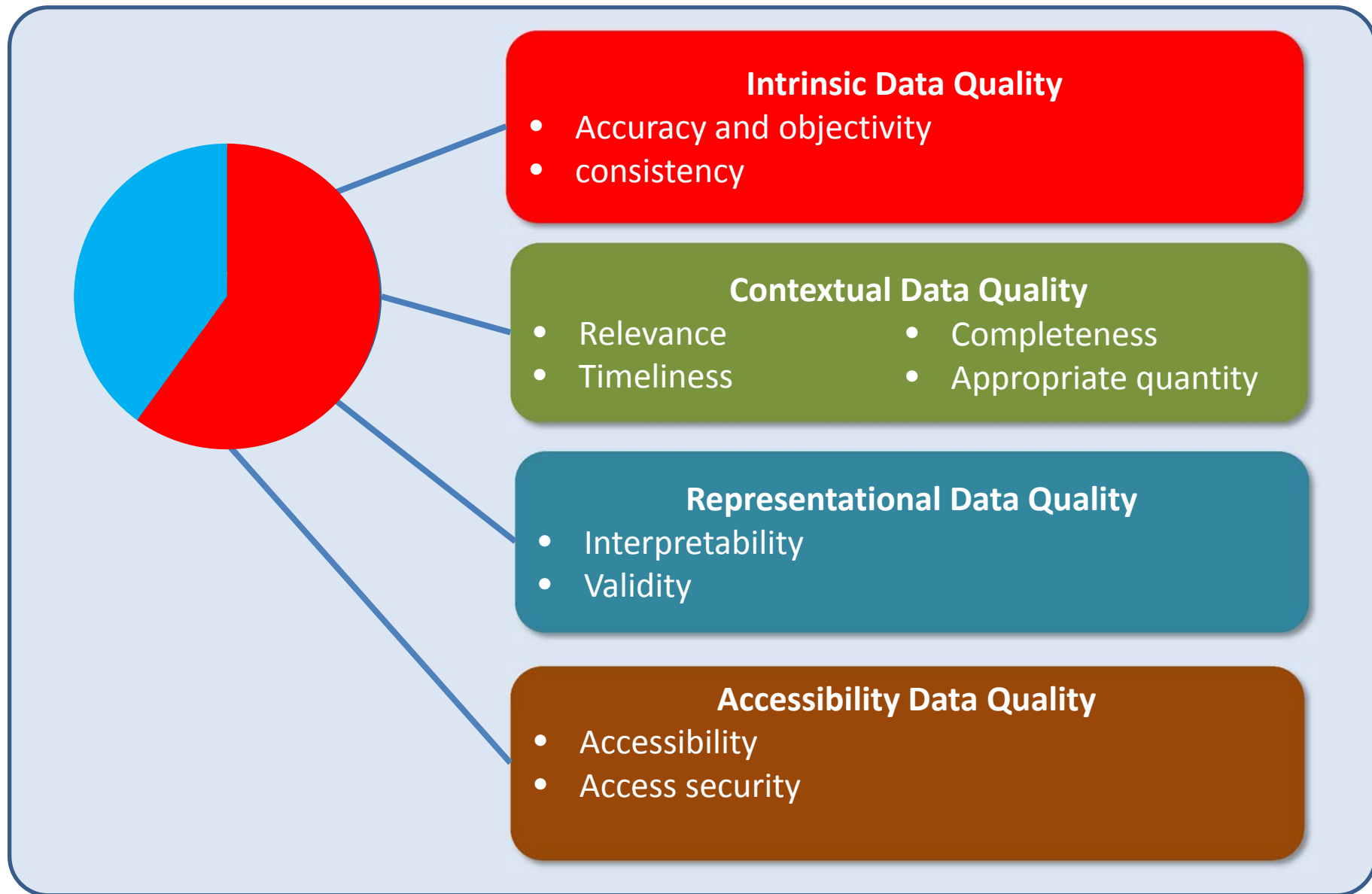
The Data Warehousing Institute (TDWI) estimates that poor quality customer data costs U.S. businesses a staggering \$611 billion a year.

Data Quality – an issue for everyone!



A recent industry survey among 500 medium-size corporations with annual sales of more than \$20 million has identified that more than 60 percent of the surveyed firms have problems with data quality.

Information Governance



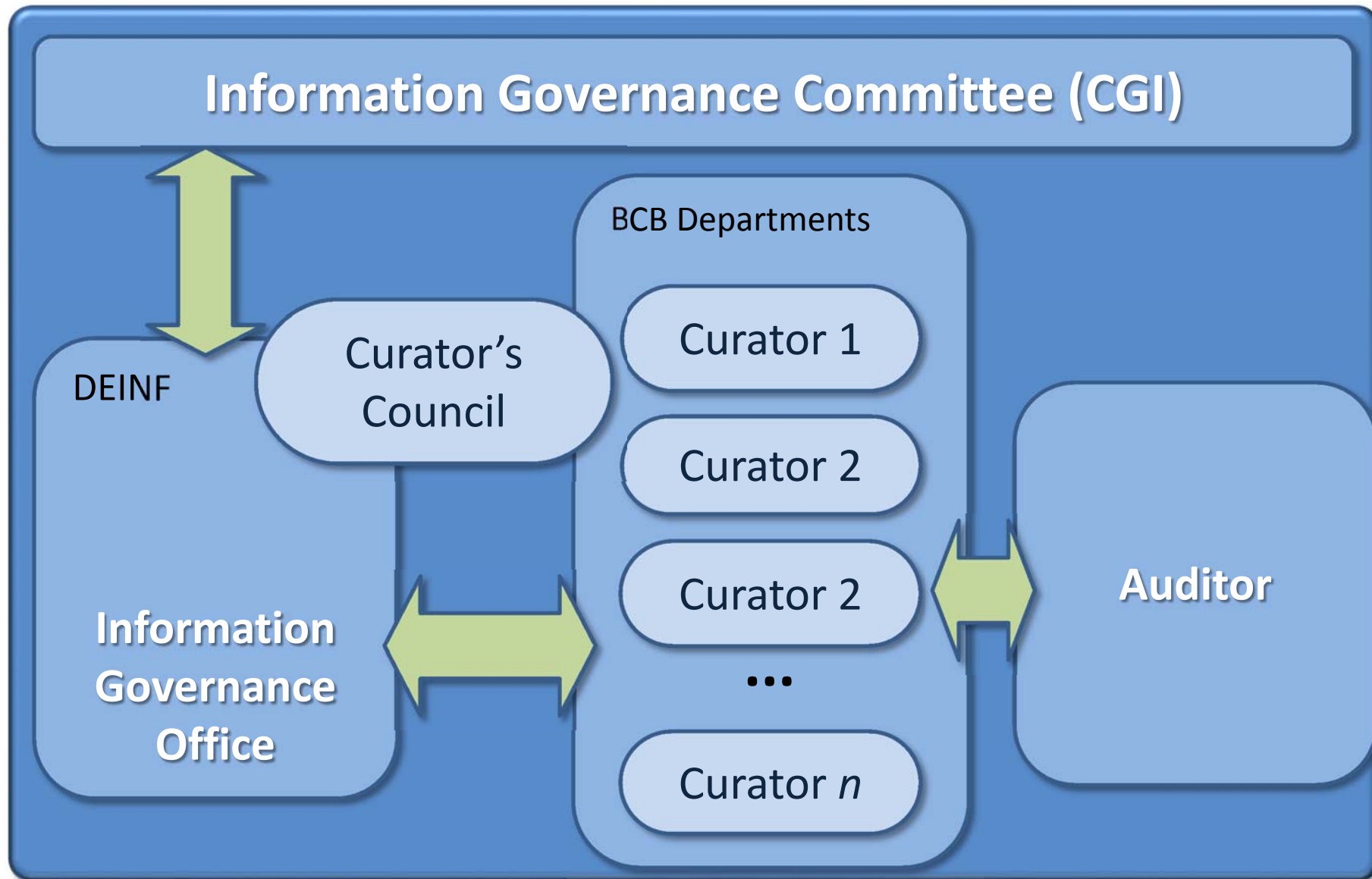
Information Governance Committee (CGI)

Created in 2013

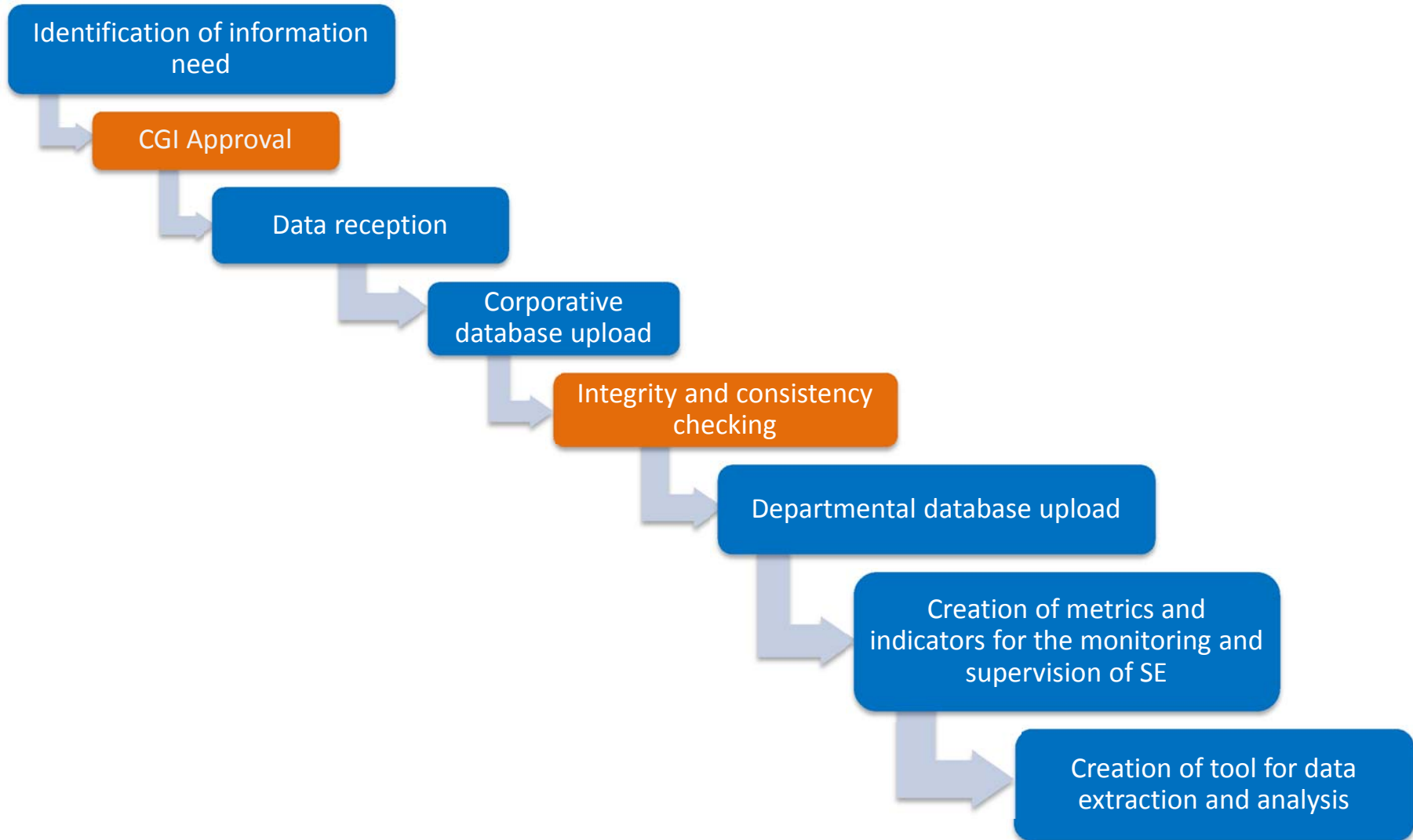
Benchmarks:

- Costs reduction
- Articulation between the creation and the use of information
- Deactivation and disposal of useless data (529 DB)
- Exemption of outdated reports to BCB
- Master data (*datos maestros*) management
- Creation of the Financial Entities Corporative Register
- Data exchange through information integration

Information Governance at BCB

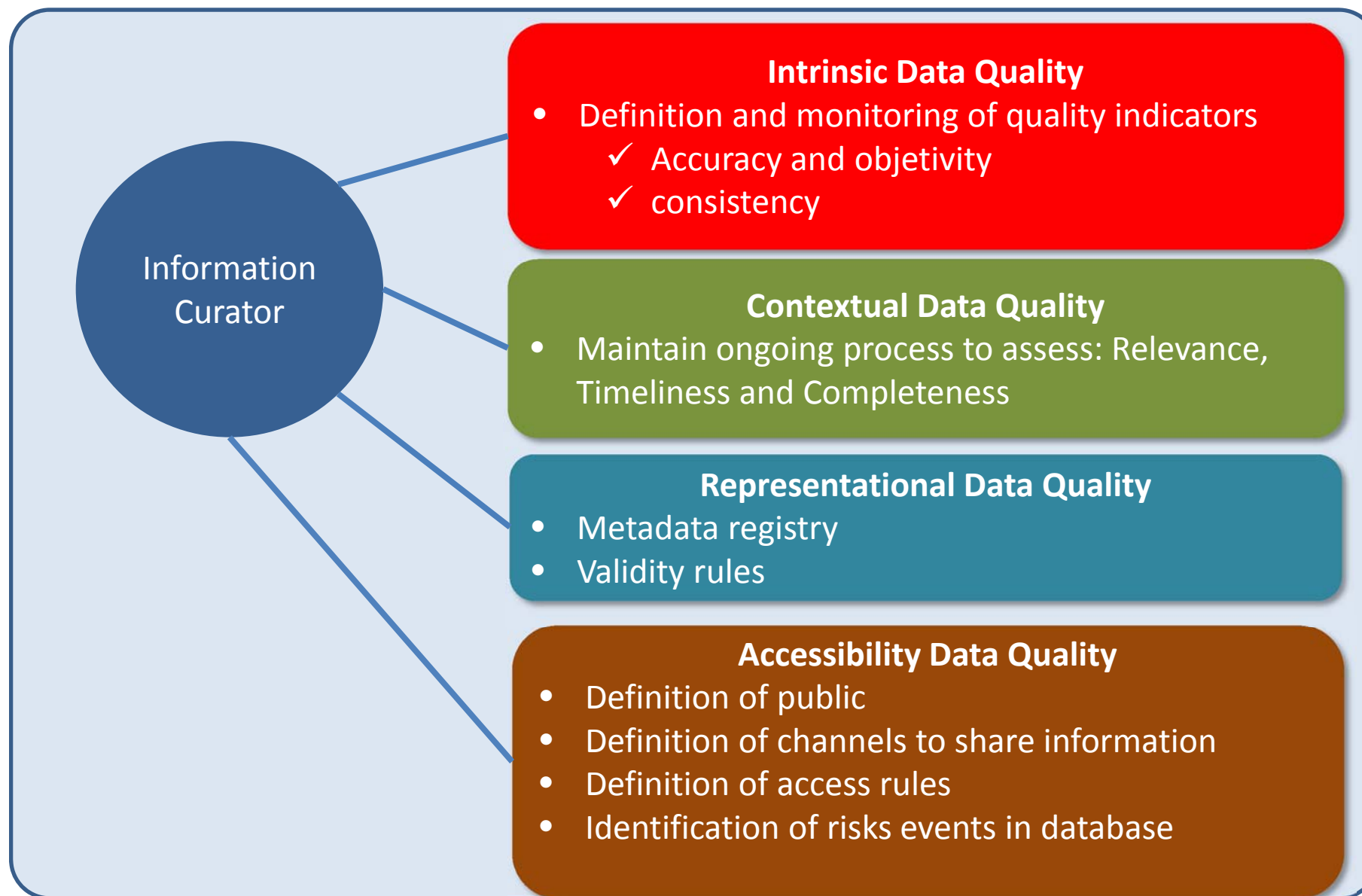


Information Governance Model

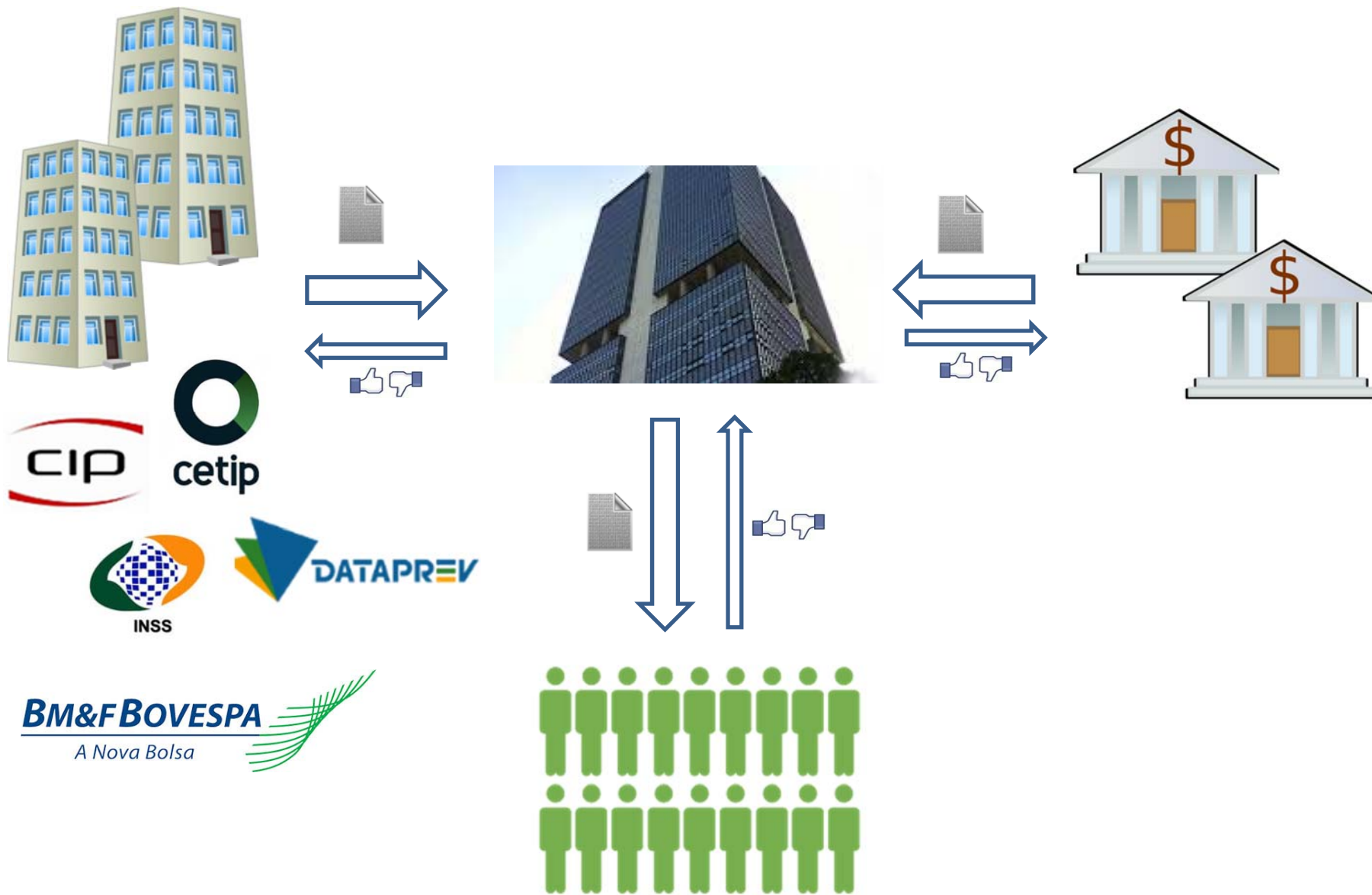


CGI defines who is responsible for the Data Quality information

Information Governance at BCB



IDQ - Quality Control must control all sources of information



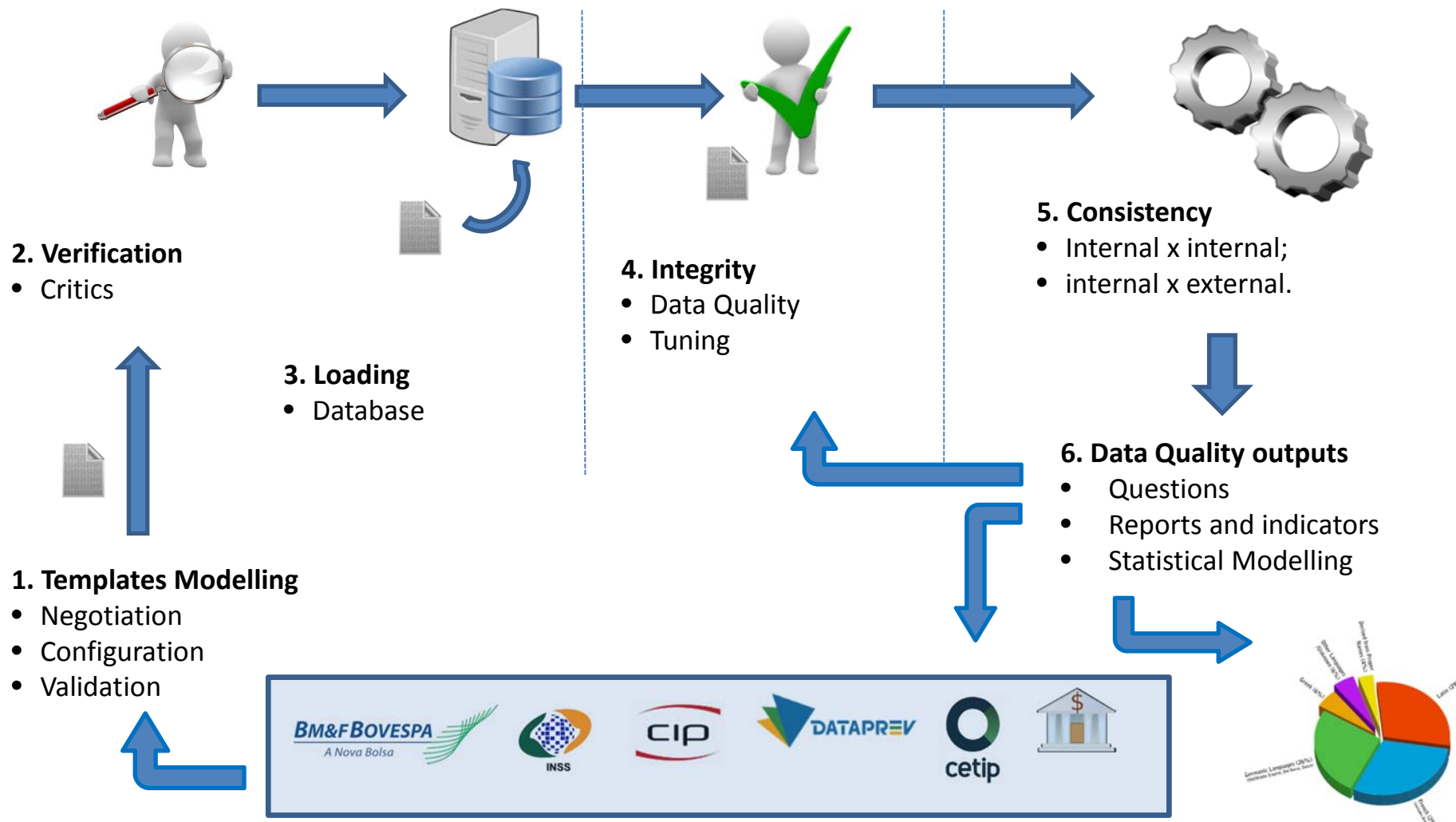
IDQ - Overview of the Integrity and Consistency Process

Departmental Governance

Data reception

Integrity

Consistency



Intrinsic Data Quality (IDQ)

Channel and format standardization

- 1 Format → XML
- 1 Channel → STA

Negotiation and Template's Disclosure

- Negotiate new acquisitions with the SE
- Template's disclosure
 - ✓ layout and instructions
 - ✓ FAQ
 - ✓ Examples

The pre-processing

- Creating pre-processing rules
- Creating xsd or other type of validator software

IDQ – Data integrity process

- **The information curator: defines the rules**
- **the TIC departament is responsible for the implementation**

- **Source Integrity**
 - ✓ Every file is encrypted and digitally signed
- **Contents Integrity**
 - ✓ Syntax validation (e.g. xsd)
 - ✓ pre-processing rules validation

In case of error, the entity must replace the document.



IDQ – Data Integrity Process

Implemented using:

- Informatica Analytics
- ETL process (scripts in PowerCenter)

Basic tests:

- Verify the presence of required information
- verify non permitted values
- verify the values Range
- verify the data format
- verify the existence of nulls
- Cross – checking between columns
- Unicity of a key

In case of error:

- The entity must replace the document or adjust the data (SCR)

Out of process:

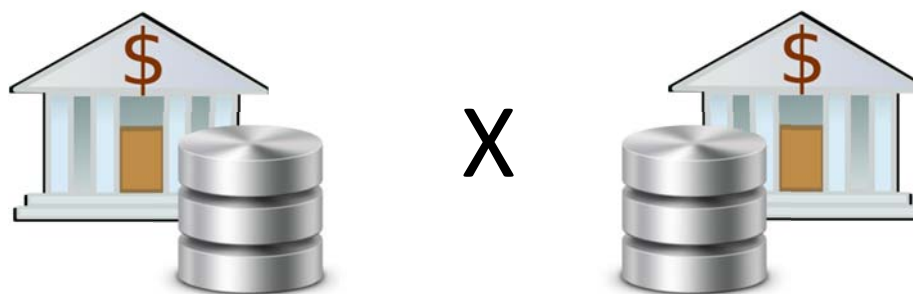
- Reports of integrity



IDQ – Internal Consistency Process

1) Cross - Checking

Compare the same data informed by the SE in different documents.



Implemented using:

- Informatica Analytics
- SQL Scripts developed by core departments
- PowerCenter scripts developed by TIC department

Beyond the quality control, this process may identify:

- Internal processes issues
- Internal inconsistencies on TI processes

E.G.

Balance Sheet x Credit Information

Balance Sheet x Operational Limits Report

2) Statistics Modelling

The most used techniques are:

- trend Analysis to find outliers and common range
- Peer group analysis with clustering technics
- Bendford's law technics

Implemented using:

- Informatica Analytics
- Software R
- SAS
- PowerCenter scripts developed by TIC department

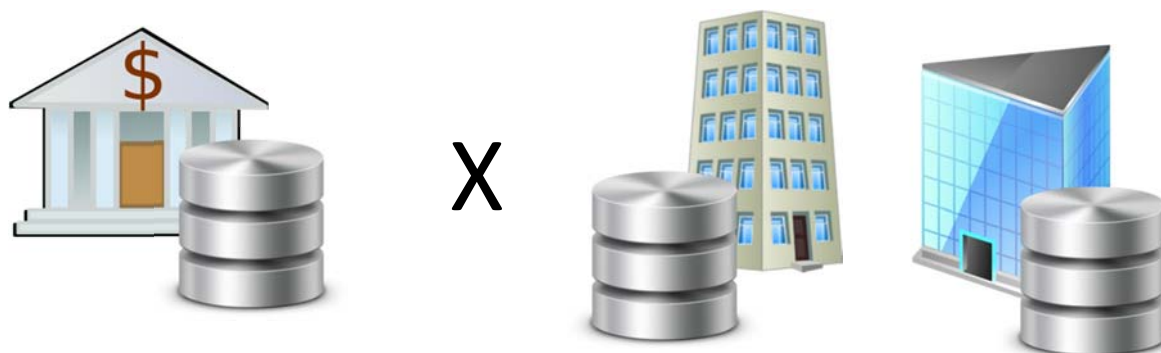
3) Disclosure of information

- Internal and external public

IDQ – External Consistency Process

External Consistency

Compare the same data informed by different sources.



Implemented by

- SQL Scripts developed by core departments
- PowerCenter scripts developed by TIC department

Beyond the quality control this process may identify:

- Internal process issues
- Illegal operations
- Irregularities

E.G.

Obituaries data x Credit Information

vehicle registers x Credit Information

Government Employee register x credit information

Integrity Process – Main Issues

Information Governance

Information reception

- Difficulties to establish agreements for data sharing among government entities
- The implementation of loading process takes a long time
- Highly bureaucratic process to acquire new information

Data Quality

- The business is much more complex than the models developed to represent them
- Detecting quality issues using internal consistency process is very complex
- Usually, it is necessary to require more information to improve data quality

Consistency and reports

- Lack of unique keys in several registers complicates the establishment of relations among them. Databases need remodeling and keys standardization
- Quality process takes long time and needs a robust IT infrastructure

Results

002B_SCRxSNG_Integridade...

http://ssrsdptprod/Reports/Pages/Report.as

Página Inicial > DESIG > REPBEH > SCRxsNG > 002B_SCRxSNG_Integridade_ChassisDuplicados

Página Inicial | Minhas Assinaturas | Ajuda

1 de 1 100%

Localizar | Próximo

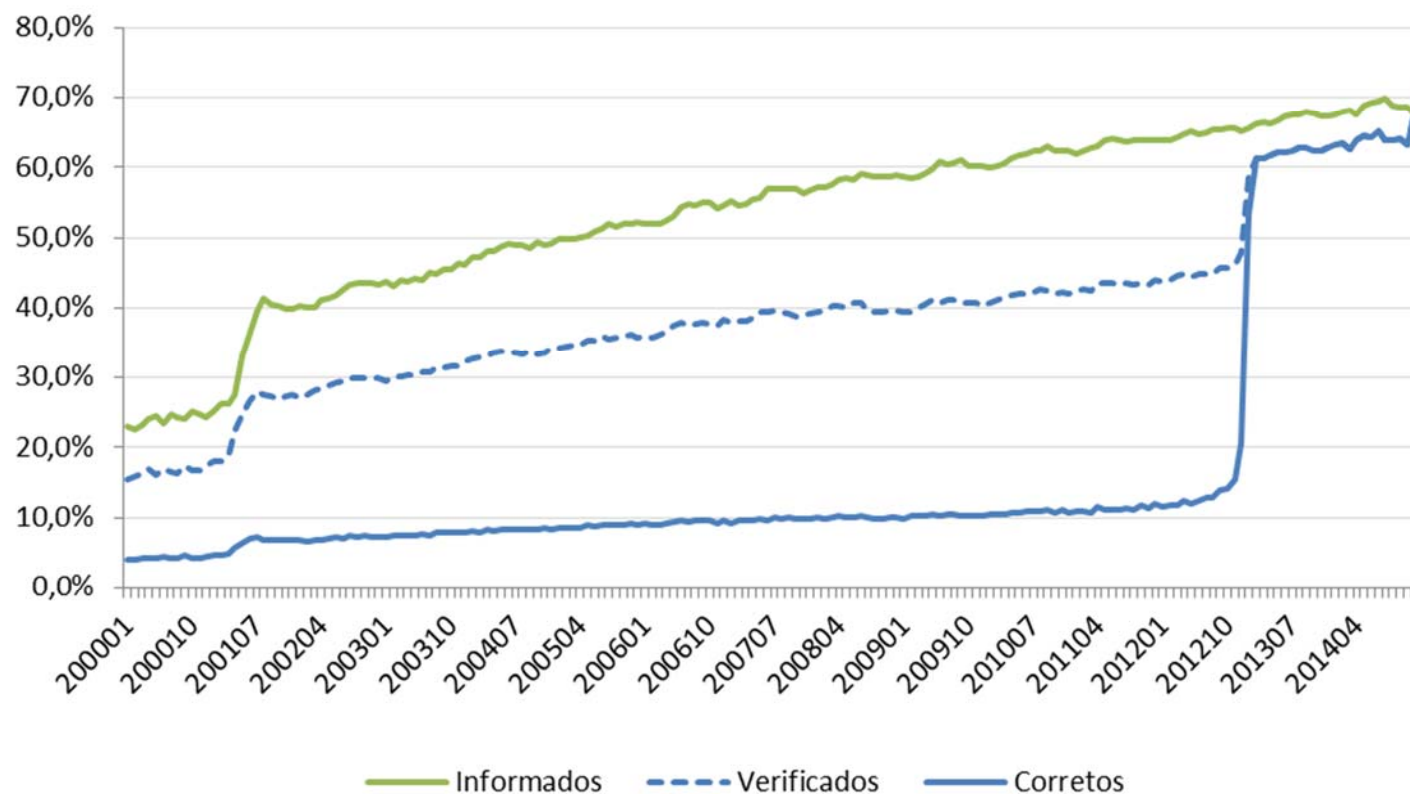
SCRxSNG: Qualidade da informação no SCR - Chassis Duplicados - Data: 201601 - Tipo de Cliente: Todos - Natureza da Operação: -1: Todas
 Tipo de veículo: Veículos Leves - Estado do veículo: Todos - Autofinanciamento: Todos

IF			Chassis duplicados em naturezas próprias						Chassis duplicados: NOP próprias IF A x cessão IF <-> A						Relevância: total duplicado como % CA (liq. prov.)	
			Número de chassis		Valor dos chassis - R\$m				Número de chassis		Valor dos chassis - R\$m				Total	Veículos
Nome	CNPJ	Cód. Congl.	Total	s/Repetição	Carteira Ativa	Inad.	Liq. Prov.	Total	s/Repetição	Carteira Ativa	Inad.	Liq. Prov.	Total	Veículos		
SFN			19.222	8.535	231,39	13,39%	198,89	1.401	698	23,49	13,85%	19,87	0,01%			
CAIXA ECONOMICA FEDERAL	00.360.305		1.558	1.503	36,61	24,55%	29,99	58	58	1,26	8,47%	1,14	0,00%			
AYMORE CFI S.A.	07.707.650	30379	1.238	1.195	27,54	2,39%	26,96	58	58	1,19	1,86%	1,17	0,12%			
BV FINANCEIRA S.A. CFI	01.149.953	51011	1.056	905	19,62	8,30%	17,54	566	562	11,08	0,25%	10,89	0,09%			
BCO DO BRASIL S.A.	00.000.000	49906	799	674	18,89	6,53%	17,38	67	67	1,16	7,07%	0,93	0,00%			
BCO BRADESCO FINANC. S.A.	07.207.996	10045	441	441	9,99	3,99%	9,55	11	11	0,11	0,00%	0,11	0,03%			
BCO BRADESCO S.A.	60.746.948	10045	540	540	9,67	7,53%	8,69	55	55	0,73	23,00%	0,56	0,00%			
BCO ITAUCARD S.A.	17.192.451	10069	704	663	9,42	9,37%	8,61	101	101	2,22	96,21%	0,09	0,01%			
BCO SANTANDER (BRASIL) S.	90.400.888	30379	500	492	9,29	4,07%	8,60	52	52	0,69	0,80%	0,65	0,00%			
BCO MERCEDES-BENZ S.A.	60.814.191	51262	191	96	6,31	7,39%	5,78	-	-	0,00	-	0,00	0,06%			
ITAÚ UNIBANCO BM S.A.	60.701.190	10069	961	592	7,93	33,87%	5,29	7	7	0,07	0,00%	0,07	0,00%			
BCO VOLKSWAGEN S.A.	59.109.165		436	426	5,97	32,31%	4,08	96	96	1,09	46,77%	0,59	0,03%			
OMNI SA CFI	92.228.410	50940	3.152	1.340	7,11	24,31%	3,55	25	24	0,31	17,76%	0,25	0,21%			
BANCO PAN	59.285.411	31323	556	553	3,35	11,83%	3,08	6	6	0,07	21,27%	0,06	0,02%			
HSBC BANK BRASIL SA BCO M	01.701.201	51152	172	164	2,83	0,40%	2,73	25	25	0,20	18,76%	0,16	0,00%			

Results – Sharing benefits

In 2012, BCB informed the INSS about quality issues in the obituary register system and suggested improvements to the quality process. After implementation, the quality of information has significantly improved.

Participação dos CPFs no total de Óbitos



Thanks

Contacts:

Andrei Vanderlei

E-mail: andrei.vanderlei@bcb.gov.br

Tel.: +55 61 3414-1834

Paula Oliveira

E-mail: paula.cristina@bcb.gov.br

Tel.: +55 61 3414-3299

Mercado de Credito

- Recibe mensualmente de **1.500 IF**:
- Operaciones de crédito de **64 millones clientes**.
- **480 millones de operaciones por mes**
- **20 mil millones de operaciones gravadas** desde 2004.

Mercado de Cambio

- **207 instituciones**
- **31.000 operaciones por día**
- **7.800.000 operaciones primarias en 2014**
- 70.000 registros diarios

Mercado de capitales, dinero y derivados

- Datos de: Selic, CETIP, BM&F Bovespa, SPB and otros
- **1 mil millones de registros por mes**
- **Mas de 900 docs procesados** cada mes (DRL y DRM).

Informaciones Contables

- recibe mensualmente **1.136 docs contables e 2.267 informes de limites operacionales**, cada trimestre **mas de 7.300 docs**, representando **mas de 2.2 millones de entradas**

Consórcio

- **11 millones de participantes** distribuidos en **20.000 grupos** y datos de **9 millones de clientes**
- Recibe **787 millones de registros** cada trimestre

Sistema de Pago

- **4.5 millones de operaciones** mes