



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

# » The use of electronic invoice data in COVID time

Almiro Moreira | António Portugal | Bruno Lima | **João Poças** «  
Jorge Magalhães | Paula Cruz | Salvador Gil | Sofia Rodrigues

UNECE Expert Meeting on Statistical Data Collection  
27-30 september 2021



# Summary

1. The Covid impact on response rates
2. Data collection during Covid: threat or opportunity?
3. Improving data quality
4. Bi-directional communication
5. Conclusions (learnings)



# 1. The Covid impact on response rates

- Response rates on **monthly** surveys \*

Monthly Surveys	Response rates (%)				
	2019	2020			
		March	April	May	June
<b>INTRASTAT</b>	<b>80,5</b>	<b>73,1</b>	75,4	75,5	77,9
Qualitative - Trade	93,5	89,7	85,2	<b>80,4</b>	87,4
Qualitative - Industry	92,5	88,2	<b>81,1</b>	<b>75,3</b>	83,4
Qualitative - Services	92,5	89,5	83,5	<b>79,3</b>	86,0
Short-Term business Statistics - Trade	79,0	77,0	<b>72,0</b>	73,0	77,0
Short-Term business Statistics - Industry	84,0	80,0	80,0	81,0	82,0
Short-Term business Statistics - Services	85,0	83,0	82,0	82,0	83,0
Index Prices on products	88,0	77,0	81,0	<b>78,0</b>	81,0

# 1. The Covid impact on response rates

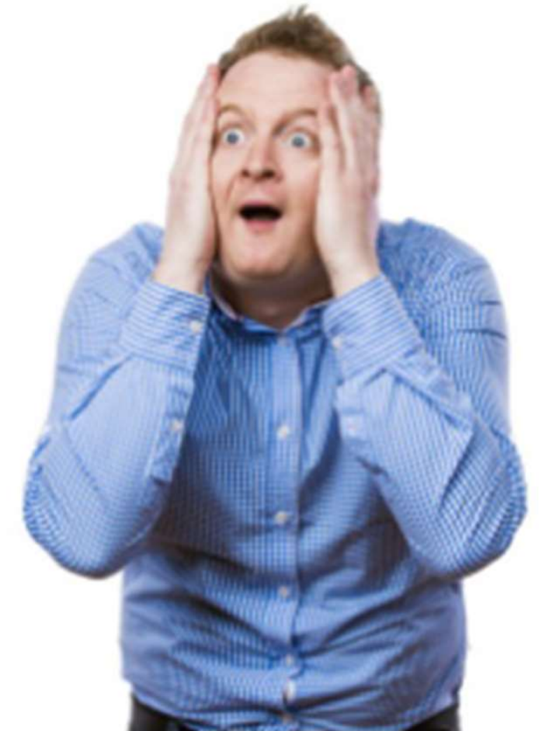
- Response rates on **annual surveys**

Annual Surveys	Response rates (%)		
	2019	2020	Diff
Services provided to enterprises	88,7	82,8	-5,9
Trade	88,3	79,3	-9,0
ICT Survey	92,5	84,0	-8,5

# 1. The Covid impact on response rates

- Not a surprise but **simply a fact** we have to deal with:

The COVID-19 pandemic has, also, dropped the response rates to business surveys, in particular during the second quarter of 2020.



# 1. The Covid impact on response rates

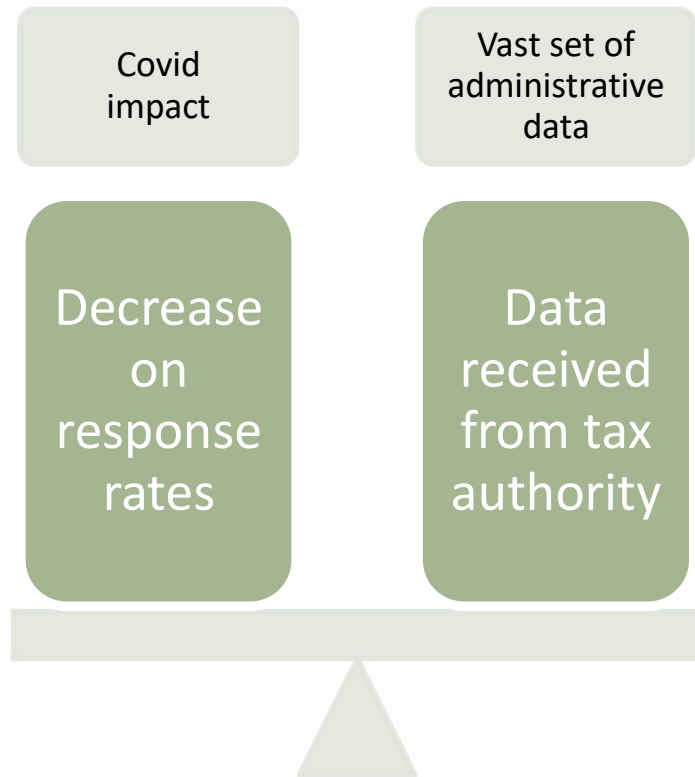
- We had two options:
  - (1) cross our fingers and hope everything was going to be back to normal or...
  - (2) **use administrative data** more intensively!
  
- In this context, information from the e-invoice system became even more relevant, filling in missing responses to the STS, and contributing to the consistency of the results obtained in the production of statistical indicators



## 2. Data collection during Covid: threat or opportunity?

- The focus on the use of administrative data has been a constant throughout the years at Statistics Portugal (INE), aiming a significant impact on the reduction of the statistical burden
- The beginning of 2020 brought, among several others, three novelties to INE's statistical production
  - the negative impact of the COVID-19 pandemic on the response rates to business surveys;
  - the receipt of a huge set of data from the invoicing of companies and;
  - the creation of a new unit dedicated to the collection, analysis and quality treatment of administrative data.

## 2. Data collection during Covid: threat or opportunity?



- Need (opportunity) to quickly be able to use a vast set of data that could suppress the lack of response and thus continue to support statistical production.
- Huge challenge: process, analyse and treat a huge amount of data received from the Tax Authority: **the e-invoice system**



### 3. Improving data quality

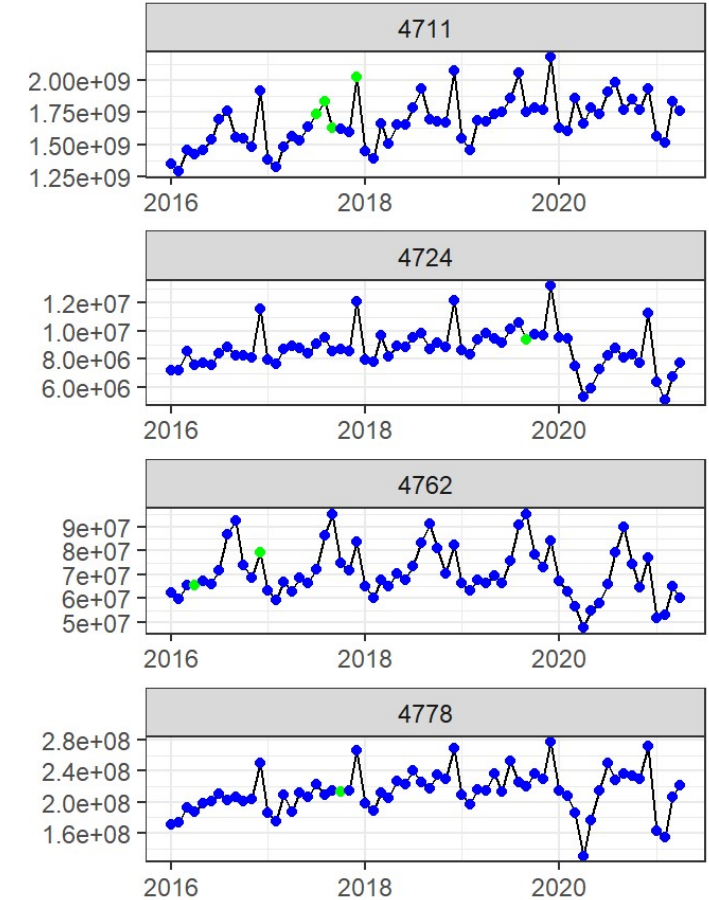
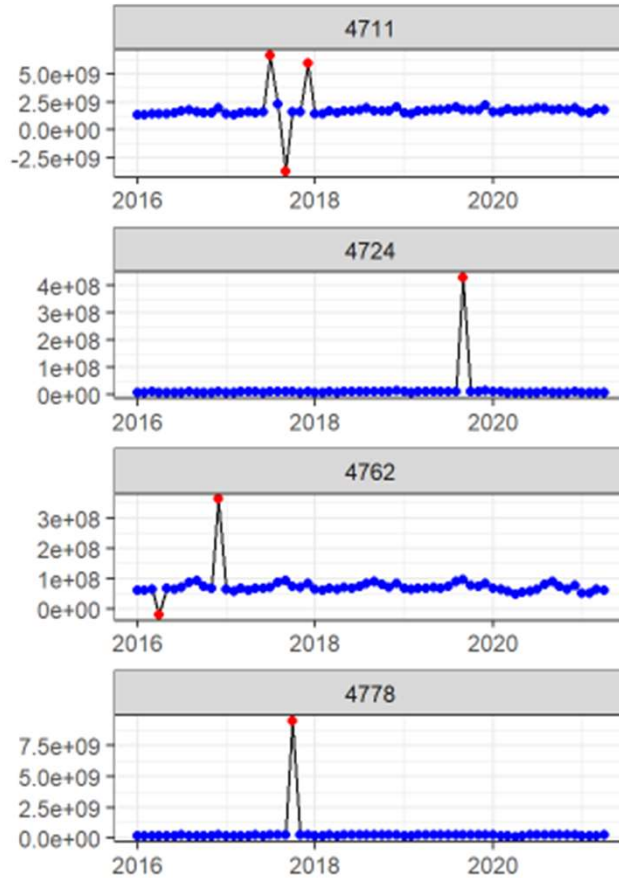
- Huge volume of data (+80 Millions records every month)
- Effort of the IT team to receive and accommodate all this data, every month  
(as of March 2020, data referenced from January 2016)
- Aware of the (potential) statistical richness of this administrative data
- Need to quickly support statistical production

## 3. Improving data quality

- Several tasks (IT team and Collection team) to process and improve data quality:
  - Validation of data structure, changes to the loading processes, verification of the number of records, validation of the fiscal identification number at the check-digit level
  - Encryption of personal identifiers;
  - Normalization of attributes (country codes);
  - Testing for consistency and comparison with other data sets;
  - Identification of outliers, elimination (1<sup>st</sup> moment) and imputation (2<sup>nd</sup> moment);
  - Classification (NACE code, type of Purchaser, ...) of entities (either Issuers or Purchasers), according to the reference date;
  - Define and provide metadata

# 3. Improving data quality

- Improving data quality



### 3. Improving data quality

- And moving from this raw data \*:

YEAR	MONTH	ISSUER	PURCHASER	COUNTRY	VALUE (€)
2021	01	501345648	500448469	PT	42,0
2021	01	579631456	999999990	PT	1 516,7
2021	01	556447988	999999990	PT	355,0
2021	01	503035649	999999990	PT	3,8
2021	01	501588971	510763375	PT	140,4
2021	01	502655984	510342175	PT	64,7
...	...	...	...	...	...
2021	01	557987887	999999990	PT	3,8

\* The data presented on this table are fictitious

### 3. Improving data quality

- Mainly (but not only) to this final data set (increasing its initial value) \*:

YEAR	MONTH	ISSUER	PURCHASER	VALUE	VALUE _TYPE	ISSUER _TYPE	PURCH _MARKET	COUNTRY	ISSUER _NUTS	ISSUER _NACE	ISSUER _SRC	PURCHASER _CLASS	...
2019	01	544556789	708683053	311,9	O	2	1	PT	170	84113	SE	CDE	...
2019	01	546999878	709406770	35	O	1	1	PT	11A	69200	SE	CDE	...
2019	01	500556066	706989139	15	O	1	1	PT	16E	69102	SE	CDE	...
2019	01	502664989	705636127	3,2	O	2	1	PT	11D	84113	SE	CDE	...
2019	01	502002001	702120944	15,94	O	2	1	PT	11A	68321	SE	CDE	...
2019	01	502002001	707386560	130	O	2	1	PT	16I	69200	SE	CDE	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2019	01	501121528	704662507	12,73	O	2	1	PT	16E	84113	SE	CDE	...

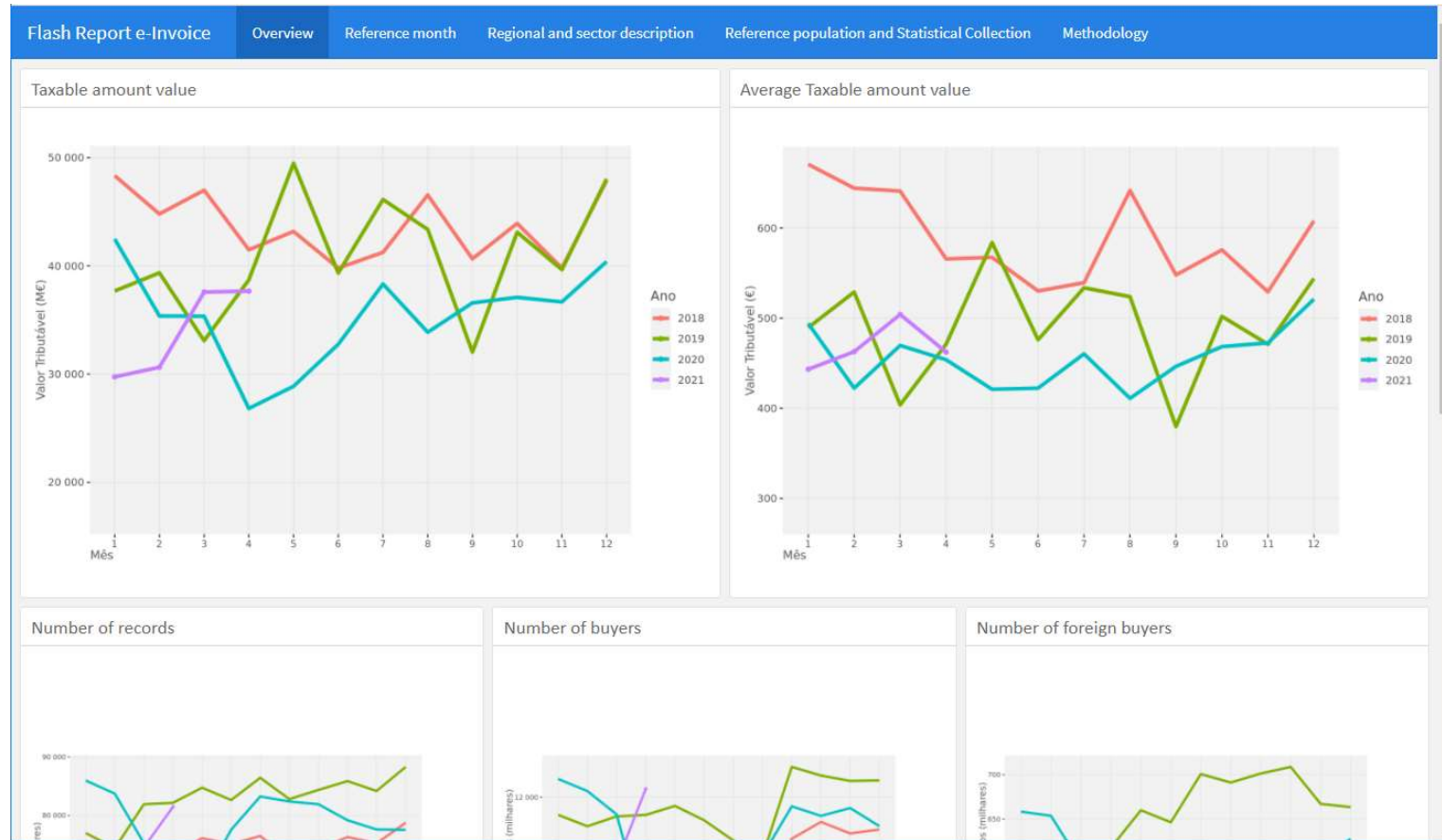
\* The data presented on this table are fictitious

## 4. Bi-directional communication

- Using administrative data for Statistical purposes should **not be seen as a one-way** communication process
- In order to promote the use of data, it is important to know the needs and expectations of its recipients in the statistical production process
- To this end, **a close dialogue with data users** was promoted in order to consider and harmonize their needs in the adoption of a data treatment and an output that would be accepted by all
- To facilitate data analysis and exploration a monthly Flash report was developed in “R Flexdashboard”

# 4. Bi-directional communication

- Example of a Flash-report about e-invoice data produced every month



## 5. Conclusions (learnings)

- The use of e-invoice data has proved to be an opportunity to strengthen the procedures for processing and analysing administrative data:
  - Was recognized as the right way to go for other sources as well
  - Contributes to the construction and fulfilment of the objectives of the National Data Infrastructure
- Investment in the acquisition of new skills, tools and techniques (data science domain), in order to overcome the difficulties of processing a massive set of data (in a very short period of time)



## 5. Conclusions (learnings)

- Strong collaboration/dialogue between different areas of the traditional statistical production process played a very important role
- In the end, the worst period of the Covid-19 pandemic can also be seen as a **catalyzer or a booster** for the **treatment and use of e-invoice data** for statistical purposes.





INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

# » The use of electronic invoice data in COVID time

Obrigado | Thank You «

UNECE Expert Meeting on Statistical Data Collection  
27-30 september 2021

