INEXDA

The Granular Data Network

INEXDA WORKING GROUP

MODES OF DATA PROVISION

Report

Results of the survey carried out in January 2021

At the 7th INEXDA meeting of July 2020, its members agreed on the launch of a Working Group on Modes of Data Provision to external researchers for research purposes.

The objective of the Working Group was to identify common features and differences and assess a potential harmonization of access procedures among INEXDA members and establish best practices.

As a first step, the Working Group launched a Survey to take stock of existing modes of data provision used by INEXDA members. In particular, the Working Group explored how data are provided to external researchers from a technical to an administrative and organizational point of view.

The survey took place between December 2020 and January 2021. Eight countries participated, together with European Central Bank and Eurostat. There were forty-four questions in this survey, organized around five groups of questions.

The following report summarizes the results of the survey. It provides a comprehensive picture as of January 2021. An update of the survey will be carried out in 2022 to capture how the pandemic has affected the provision of data to researchers.

The main takeaways are:

1. Most respondents provide access to data in several ways, often combining the provision of public use files and onsite solutions;
2. Access modes differ with respect to content and/or degree of anonymization;
3. The number of researchers per project averages two;
4. Data preparation is usually done for standardized datasets (sometimes combined with custom-datasets);
5. Respondents other than those providing public files use some form of confidentiality agreements;
6. Output is usually checked manually;
7. There are different approaches to governance;

GROUP 1 – Access modes

The first group of questions aimed at identifying the different access modes across respondents, i.e. how an external researcher gets access to data (e.g. on-site access, remote access). Access modes may differ depending on a specific type of researcher conducting a project, a specific type of research field or a specific dataset family.[1]

Across the 10 respondents, the survey identified 24 access modes, 3 being set to open in 2021 or 2022. See Annex 1 and 2 for more information.

Table 1.1 – Number of access modes across respondents

Country	Institution	Number of access modes
DE	Deutsche Bundesbank	5
ES	Banco de España	2[2]
EU	European Central Bank	3
EU	Eurostat	2
FR	Banque de France	2
IT	Banca d'Italia	2
MX	Banco de México	2
PT	Banco de Portugal	3
RU	Bank of Russia	1
TR	Central Bank of the Republic of Turkey	2

---

[1] See INEXDA Annodata Schema, a common taxonomy describing all legal, technical, and organizational rules that have to be considered when making micro data available for researchers.

[2] A third access mode has been created (BELab) between the launch of the survey and the publication of this report. Meanwhile new features are being implemented.

Among them, 5 offer multiple ways to access data.[3] On-site access (both secure or not secure) is the most frequent type of access among respondents with 13, while only 4 institutions offer a remote access to data for researchers. 7 access modes consist of the provision of scientific or public use file (download).

France (ODR2) is the sole institution charging fees to researchers.

Table 1.2 – Access modes in detail (See Annex 2 for more information)

Access modes	Secure on-site access	On-site access	Remote execution	Remote access	Download	Other
DE - PUFs					o	
DE - B-Data	o					
DE - USTAN/JANIS	o					
DE - S-Data	o		o			
DE - SUFs					o	
ES - CAER	o			o		
ES - EFF - PUF					o	
ECB - ESRB (EMIR bridge)		o				
ECB – Astra – external researchers					o	o
ECB – Darwin – ESCB researchers					o	o
Eurostat - SciUF					o	
Eurostat - SecUF	o					
FR - ODR	o					
FR - ODR2				o		
IT - BIRD			o			
IT - PUF					o	
MX - EconLab ER	o					
MX - LISF	o					
PT - ESA			o	o		
PT - ISA		o				
PT - SA			o			
RU - DRER	o					
TR - os		o				
TR - scr_os	o					

---

[3] **Secure on-site access**: Provision in the premises of the institution in a dedicated secure environment (data room). **On-site access**: Provision via the institution's IT infrastructure. **Remote execution**: Researcher sends code and RDC sends back results. **Remote on-site access**: Remote access for internal researchers (e.g. teleworking). **Remote access**: Researcher can access data stored in another location remotely from own institution in a controlled environment. **Download**: Files available for download and copy used according to the terms of use (e.g. scientific use files or public use files).

The number of users and projects varies greatly across access modes, but the number of researchers by project average below or around 2 in most cases. Provision of scientific or public use files naturally capturing more researchers in total.

Table 1.3 – Number of projects, researchers and internal resources

Access modes	Projects	Researchers	Average		Internal resources (FTE)
DE - PUFs					
DE - B-Data	76	204	2.68		2
DE - USTAN/JANIS	52	96	1.85		1
DE - S-Data	216	469	2.17		3
DE - SUFs	277	453	1.64		1
ES - CAER	8	18	1.87		2
ES - PUF		2543			
ECB - ESRB (EMIR bridge)	6	18	3.00		
ECB - Astra		2050			3
ECB - Darwin		75			3
Eurostat - SciUF	940	2000	2.13		4
Eurostat - SecUF	32	60	1.88		5
FR - ODR	40	60	1.50		2
FR - ODR2					4
IT - BIRD	130	130	1.00		0.25
IT - PUF					0.15
MX - EconLab ER	40	15	0.38		2
MX - LISF					3
PT - ESA	65	93	1.43		6.5
PT - ISA	32	33	1.03		6.5
PT - SA	4	4	1.00		6
RU - DRER	3	5	1.67		5
TR - os	5				2
TR - scr_os		45			6

In terms of researchers, access is primarily granted to Universities or Research institutions. 'Financial or Commercial Firms' can solely access data through public use files.

*Table 1.4 – Researchers origins*

Access modes	Universities	Research institutions	Individual researchers	Other national institutions	International institutions	Other central banks	Financial institutions	Other commercial firms
DE - PUFs	o	o	o	o	o	o	o	o
DE - B-Data	o	o		o	o	o		
DE - USTAN/JANIS	o	o		o	o	o		
DE - S-Data	o	o		o	o	o		
DE - SUFs	o	o		o	o	o		
ES - CAER	o	o		o	o	o		
ES – EFF - PUF			o					
ECB - ESRB (EMIR bridge)	o	o	o			o		
ECB - Astra	o	o	o	o	o		o	
ECB - Darwin						o		
Eurostat - SciUF	o	o						
Eurostat - SecUF	o	o						
FR - ODR	o	o	o					
FR - ODR2	o	o	o	o				
IT - BIRD	o	o	o	o	o	o		
IT - PUF	o	o	o	o	o	o	o	o
MX - EconLab ER	o	o		o	o	o		
MX - LISF	o	o	o	o	o	o		
PT - ESA			o					
PT - ISA			o					
PT - SA			o					
RU - DRER	tbd							
TR - os			o					
TR - scr_os	o	o	o					

GROUP 2 – DATA

The second group of questions aimed at identifying the types of data accessible to researchers within an access mode. It refers to the source of the data and the anonymization methods used to allow for sharing confidential information with researchers, if any.

In term of data source, 12 access modes offer access to data on a single thematic ('Economy and finance' or 'Industry, trade and services'), while the remaining 12 access modes offer access to multiple thematics. A large majority of access modes (23) provide access to data on the theme 'Economy and finance'. Zooming on 'Economy and finance', data on 'Firms and Banks' are the most often made available to researchers.

Chart 2.1 – Data source by theme
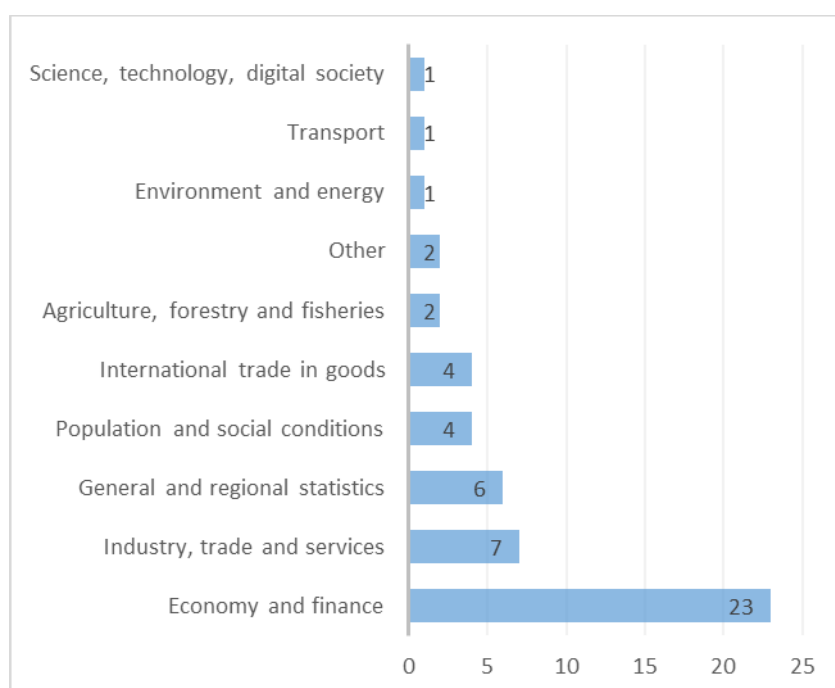
Chart 2.2 – 'Economy and finance' by Sector *Chart 2.3 – 'Economy and finance' by Type*
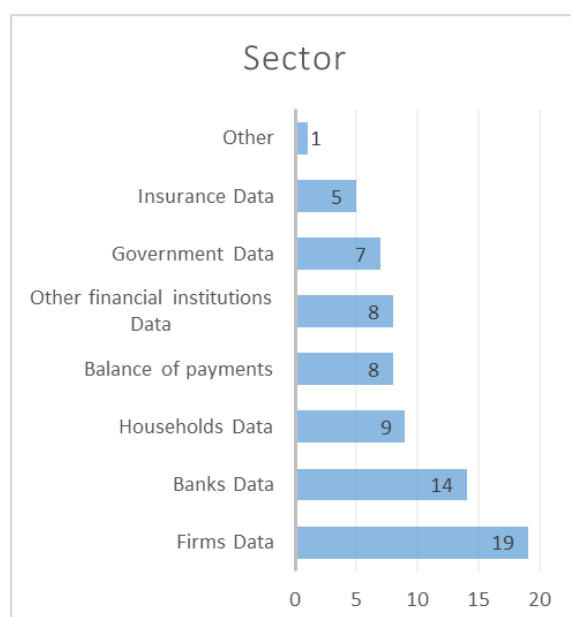
6

Table 2.1 – Data source by theme in detail

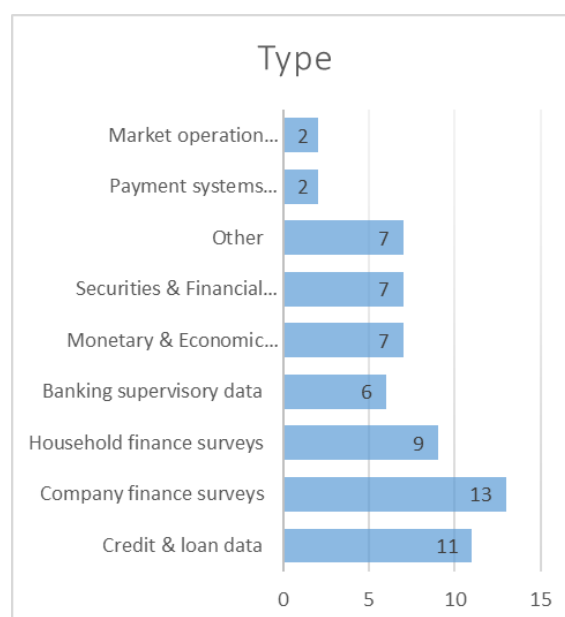| Access modes	Economy and finance	Industry, trade and services	General and regional statistics	Population and social conditions	International trade in goods	Agriculture, forestry and fisheries	Other	Environment and energy	Transport	Science, technology, digital society
DE - PUFs	o	o	o		o					
DE - B-Data	o									
DE - USTAN/JANIS	o									
DE - S-Data	o	o			o					
DE - SUFs	o			o						
ES - CAER	o									
ES - EFF - PUF	o									
ECB - ESRB (EMIR bridge)	o		o		o		o			
ECB - Astra	o	o								
ECB - Darwin	o	o								
Eurostat - SciUF	o	o		o		o			o	o
Eurostat - SecUF		o								
FR - ODR	o		o							
FR - ODR2	o		o							
IT - BIRD	o									
IT - PUF	o			o						
MX - EconLab ER	o	o	o	o	o	o		o		
MX - LISF	o									
PT - ESA	o									
PT - ISA	o									
PT - SA	o									
RU - DRER	o		o				o			
TR - os	o									
TR - scr_os	o									

With regards to data preparation, in a large majority of cases, 16, datasets have not been modified to specific needs. Data are said to be "Standardized", or identical across research projects, as opposed to "Customized". The remaining 8 access modes offer a combination of standardized and customized datasets.

This section concerns solely the 18 access modes providing on-site or remote access to data. Among these relevant cases, 15 access modes allow researchers to bring their own data to the environment. 1 institution (MX – EconLab IR) allows for an automatic import of external data while 2 institutions (MX – LISF and TR – os) do not allow for any import.

Table 2.2 – Data preparation and import of external data

Access modes	Data preparation		Data import
DE - PUFs	SD		-
DE - B-Data	CM		Y
DE - USTAN/JANIS	SD		Y
DE - S-Data	SD		Y
DE - SUFs	SD		-
ES - CAER	SD		Y
ES - EFF - PUF	SD		-
ECB - ESRB (EMIR bridge)	SD		Y
ECB - Astra	SD		-
ECB - Darwin	SD		-
Eurostat - SciUF	SD		Y
Eurostat - SecUF	SD		Y
FR - ODR	SD		Y
FR - ODR2	SD		Y
IT - BIRD	SD & CM		Y
IT - PUF	SD		-
MX - EconLab ER	SD & CM		Y+
MX - LISF	SD & CM		N
PT - ESA	SD & CM		Y
PT - ISA	SD & CM		Y
PT - SA	SD & CM		Y
RU - DRER	SD & CM		Y
TR - os	SD		N
TR - scr_os	SD		Y

Notes:
SD: Standardized, CM: Custom
Y: Data import allowed, Y+: Automatic data import , N: No data import, -: Not applicable

Concerning anonymization[4], 2 access modes provide non-anonymized data to researchers. In 7 cases, various anonymization methods are used within a single access mode. Depending on the level of confidentiality of the underlying data, different strategies to create modified datasets are applied. Formal anonymization is the most commonly used method (15). Full anonymization is the preferred method for public use files.

Chart 2.4 – Anonymization methods
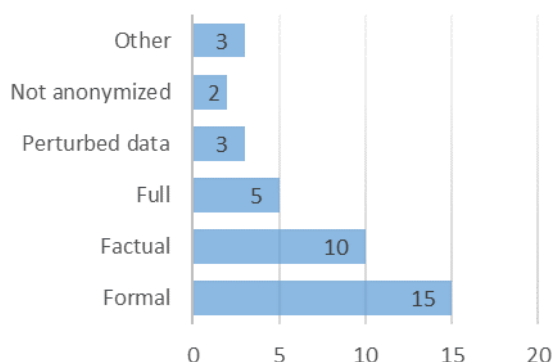
Table 2.3 – Anonymization methods in detail

Access modes	Formal	Factual	Perturbed data	Not anonymized	Other	Full
DE - PUFs						o
DE - B-Data	o					
DE - USTAN/JANIS	o					
DE - S-Data	o					
DE - SUFs		o				
ES - CAER[5]				o		
ES - EFF - PUF	o					o
ECB - ESRB (EMIR bridge)	o			o		
ECB - Astra	o					o
ECB - Darwin	o					
Eurostat - SciUF	o	o				
Eurostat - SecUF	o					
FR - ODR	o					
FR - ODR2	o					

---

[4] **Non-anonymized**: No anonymization (i.e. raw data). **Formal anonymization**: Deletion of direct identifiers such as names, addresses, and other identifier (e.g. LEI). No direct identification possible. **Factual anonymization**: Identification possible only with significant effort. **Perturbed data**: Data modified using statistical methods. **Full anonymization**: No identification possible

[5] New features are being implemented between the launch of the survey and the publication of this report.

Access modes	Formal	Factual	Perturbed data	Not anonymized	Other	Full
IT - BIRD		o				
IT - PUF						o
MX - EconLab ER	o	o				
MX - LISF	o	o				
PT - ESA	o	o	o		o	
PT - ISA	o	o	o		o	
PT - SA					o[6]	
RU - DRER	o	o	o			o
TR - os		o				
TR - scr_os		o				

Notes:
For most cases, factual anonymization is applied in addition to formal anonymization.

According to German law, formally anonymised data that is accessed in a secure environment satisfies the criteria of factual anonymization. Therefore, all external researchers who work inside the secure environment use factual anonymised data.

---

[6] Others include: (1) shuffled data (contains "shuffled" variables, that is, variables that keep their original values but have been sorted randomly and independently of the rest of the dataset to break the link for the observations; and (2) randomized data (data that is both perturbed and shuffled).

GROUP 3 – Architecture of the environment

The third group of questions aimed at identifying the type of environment provided to external researchers to analyze data, in case of on-site or remote access.

Among the 13 on-site access, the number of workstations made available to researchers range from 1 (ECB- ESRB and TR) to 16 (DE).

Table 3.1 – Environment

Access modes	Workstation	Virtual Machine	Remote access Software	Authentification
[On-site access]				
DE - B-Data	16	10	Citrix	Multi factor: Smart card
DE - USTAN/JANIS	16	10	Citrix	Multi factor: Smart card
DE - S-Data	16	10	Citrix	Multi factor: Smart card
ECB - ESRB (EMIR bridge)	1			Single factor
Eurostat - SecUF	2			Other
FR - ODR	4		Citrix	Single factor
MX - EconLab ER	3			
MX - LISF	3		Citrix	Multi factor: RSA Virtual token
PT - ISA	10			Single factor
RU - DRER	2	5	Citrix	Single factor
TR - os	2			
TR - scr_os	1	4	Other	Single factor
ES - CAER	4	2	Citrix	Multi factor: RSA Physical token
[Remote access]				
Eurostat - SecUF			Citrix	Multi factor: RSA Physical token
FR - ODR2		70	Citrix	Multi factor: Google authenticator
PT - ESA			Other	Single factor
PT - SA				
IT - BIRD			Other	Other
[Download]				
DE - PUFs				
DE - SUFs				Other
ES - PUF				
ECB - Astra				Multi factor: Other
ECB - Darwin				Other
Eurostat - SciUF				Multi factor
IT - PUF				

Table 3.2 – Environment of the dedicated server

Access modes	CPU	Memory (Gb)	Storage (Gb)	OS	File storage system
[On-site access]					
DE - B-Data	2	192		WIN	Filesystem
DE - USTAN/JANIS	2	192		WIN	Filesystem
DE - S-Data	2	192		WIN	Filesystem
EU - ESRB (EMIR bridge)				WIN	Hadoop
EU - SecUF				WIN	
FR - ODR				WIN	Filesystem
MX - EconLab ER	14	512	4000	WIN	Filesystem
MX - LISF	8	16	1150	WIN	Microsoft SQL Server
PT - ISA	64	1007	7000	LINUX	Filesystem
RU - DRER				WIN	Oracle
TR - os				WIN	
TR - scr_os	24	512	3000	WIN	Filesystem
ES - CAER				WIN	
[Remote access]					
EU - SecUF				WIN	Filesystem
FR - ODR2				WIN	Filesystem
PT - ESA	56	440	3000	LINUX	Filesystem
IT - BIRD				LINUX	Filesystem

16 access modes offer software to researchers. R, Python and Stata are available in a large majority of cases. 3 have tools to share and version code (e.g. github)

All offer updates and additional software upon request, while 5 also provide a self-service local repository.

Chart 3.1 – Provision of software
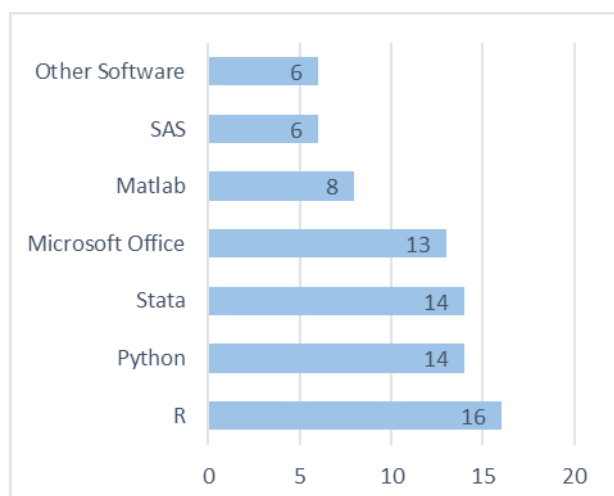
GROUP 4 – Governance

The forth group of questions aimed at identifying the governance surrounding the access of data by researchers. The governance includes the parties involved in the decision process as well as the administrative measures in place to ensure that no confidential data is ever disclosed.

Governance differs among respondents but 4 models can be identified. In 8 cases, a governance body approves the data access request. In 6 cases, the decision is fully delegated to the research data center. In 5 additional cases, the data producer and/or owner are decision makers. Finally, in 5 cases, the research data center and producers are involved in the approval process.

Table 4.1 – Data request approval

Access modes	Data Owner	Data Producer	Research Center Manager	Research Center Staff	Chief Data Officer	Governance Body
ES - CAER						o
FR - ODR						o
FR - ODR2						o
MX - EconLab ER						o
MX - LISF						o
RU - DRER						o
DE - SUFs				o		
IT - BIRD				o		
IT - PUF				o		
PT - ESA			o	o		
PT - ISA			o	o		
PT - SA			o	o		
ECB - ESRB (EMIR bridge)	o					
ECB - **Astra**	o	o				
ECB - **Darwin**	o	o				
TR - os	o	o				
TR - scr_os	o	o				
DE - B-Data		o		o		
DE - USTAN/JANIS		o		o		
DE - S-Data		o		o		
Eurostat - SciUF	o	o	o	o		
Eurostat - SecUF	o	o	o	o		
DE - PUFs	-	-	-	-	-	-
ES - EFF - PUF	-	-	-	-	-	-

Among the 6 models involving a governance body, composition differs. In most cases, data producers, owners and internal researchers are represented, while external researchers and legal advisors are represented to a lesser extent.

Table 4.2 – Governance body composition

Access modes	Data Owners	Data Producers	Chief Data Officer	Legal department	Internal researchers	External researchers
ES - CAER					o	
FR - ODR	o	o		o	o	o
FR - ODR2	o	o		o	o	o
MX - EconLab ER	o	o				
MX - LISF	o	o				
RU - DRER	o		o	o	o	

To ensure that confidential information is not disclosed, all access modes (23), other than public use files, require the signature of a confidentiality agreement by the researcher or research body.

Additionally, and in most cases (17), the output is controlled by the institution providing the data. 8 access modes prescribe methods to researchers to handle their output.

Table 4.3 – Confidential agreement and output control

Access modes	Confidentiality agreement signed by research body	Confidentiality agreement/declaration signed by researcher		Output control by researchers according to prescribed methods	Output controlled in the institution (e.g. the data provider)
DE - PUFs	-	-		-	-
DE - B-Data	o	o		o	o
DE - USTAN/JANIS		o		o	o
DE - S-Data		o		o	o
DE - SUFs		o		o	
ES - CAER		o		o	o
ES - EFF - PUF	-	-		-	-

Access modes	Confidentiality agreement signed by research body	Confidentiality agreement/declaration signed by researcher		Output control by researchers according to prescribed methods	Output controlled in the institution (e.g. the data provider)
ECB - ESRB (EMIR bridge)	o	o		o	o
ECB - Astra		o			
ECB - Darwin		o			
Eurostat - SciUF	o	o		o	
Eurostat - SecUF	o	o			o
FR - ODR		o			o
FR - ODR2		o			o
IT - BIRD		o			o
IT - PUF	-	-		-	-
MX - EconLab ER		o			o
MX - LISF		o			o
PT - ESA		o			o
PT - ISA		o			o
PT - SA		o			o
RU - DRER	o			o	o
TR - os		o			o
TR - scr_os	o				o

In term of sanctions, 17 access modes foresee action in case of misuse of the confidential information provided to researchers.

Once approved, access to data is limited in time by most access modes (20). Only 3 access modes other than public use files provide an unlimited access to data.

Table 4.4 – sanctions and duration of use

Access modes	Sanction		Duration
DE - PUFs			u
DE - B-Data	Y		l
DE - USTAN/JANIS	Y		l
DE - S-Data	Y		l
DE - SUFs	Y		l
ES - CAER	N		l
ES - EFF - PUF			u
ECB - ESRB (EMIR bridge)	Y		l
ECB - Astra	Y		l
ECB - Darwin			l

Access modes	Sanction		Duration
Eurostat - SciUF	Y		l
Eurostat - SecUF	Y		l
FR - ODR	Y		u
FR - ODR2	Y		u
IT - BIRD	N		u
IT - PUF			u
MX - EconLab ER	N		l
MX - LISF	N		l
PT - ESA	Y		l
PT - ISA	Y		l
PT - SA	Y		l
RU - DRER	Y		l
TR - os	Y		l
TR - scr_os	Y		l

Notes:
Y: Sanctions foreseen, N: No sanction foreseen
u: unlimited use of data, l: limited use of data

GROUP 5 – Results and output control

The fifth group of questions aimed at identifying how results are handled and output controlled.

This section concerns solely the 18 access modes providing on-site, secure on-site or remote access to data. In all cases, results can be exported from secure environments upon request and in aggregate form. Direct or autonomous exports are not made possible.

Relative to the output control, manual intervention is the rule. Although 6 access modes provide automatisation in addition to manual output control.

Table 5.1 – Output control

Access modes	No output control	Yes, Manual	Yes, Automatic
DE - PUFs	-	-	-
DE - B-Data		o	o
DE - USTAN/JANIS		o	o
DE - S-Data		o	o
DE - SUFs	-	-	-
ES - CAER[7]		o	
ES - EFF - PUF	-	-	-
ECB - ESRB (EMIR bridge)		o	
ECB - Astra	-	-	-
ECB - Darwin	-	-	-
Eurostat - SciUF		o	
Eurostat - SecUF		o	
FR - ODR		o	
FR - ODR2		o	
IT - BIRD		o	
IT - PUF	-	-	-
MX - EconLab ER		o	
MX - LISF		o	
PT - ESA		o	o
PT - ISA		o	o
PT - SA		o	o
RU - DRER		o	
TR - os		o	
TR - scr_os		o	

---

[7] Automatic output control tools under development

Annex 1 – Access modes by respondents

Country Institution	Name	Short name	Year of Creation of the service
DE **Deutsche Bundesbank**	Aggregated statistics	PUFs	
	Banking supervisory data	B-Data	2015
	Company data	USTAN/JANIS	2014
	Statistical Data (Standard Case)	S-Data	2014
	Survey data	SUFs	2014
ES **Banco de España**	Controlled Access for External Researchers	CAER	2019
	Public files	EFF - PUF	2002
EU **European Central Bank**	ESRB - Alberto Giovannini Programme on Data Science	ESRB (EMIR bridge)	
	ECB-ASTRA – external researchers	Astra	2020
	ECB-Darwin – ESCB researchers	Darwin	2009
EU **Eurostat**	Access to scientific use files	SciUF	2013
	Access to secure use files	SecUF	2013
FR **Banque de France**	Open Data Room	ODR	2016
	Open Data Room 2.0	ODR2	2021
IT **Banca d'Italia**	Banca d'Italia Remote access to micro Data	BIRD	2008
	Public Use File	PUF	1977
MX **Banco de México**	EconLab External Researchers	EconLab ER	2015
	Laboratorio de Información del Sistema Financiero	LISF	2021
PT **Banco de Portugal**	External Server Access	ESA	2017
	Internal Server Access	ISA	2016
	Surrogate Access	SA	2019
RU **Bank of Russia**	Data room for external researchers (currently in design)	DRER	2022
TR **Central Bank of the Republic of Turkey**	On-site access	os	
	Secure on-site access	scr_os	

Annex 2 – Access modes detailed description by respondents

- **DE** - Deutsche Bundesbank

Short name	Description
S-Data	The Bundesbank provides external scientific researchers access to formally anonymised micro data on firms, banks, insurances, balance of payments and other financial institutions. The data has been collected for statistical purposes and can be accessed in the secure environment of the Bundesbank's RDSC and, for experienced external researchers, via remote execution. Any access is subject to application and approval by the RDSC. The RDSC provides standard datasets and conducts an output and publication control before any results can leave the RDSC. According to German law, formally anonymised data that is accessed in a secure environment satisfies the criteria of factual anonymization. Therefore, all external researchers who work inside the secure environment use factual anonymised data.
USTAN/JANIS	The Bundesbank provides external scientific researchers access to formally anonymised micro data on firms. The data can be accessed in the secure environment of the Bundesbank's RDSC only. Any access is subject to application and approval by the RDSC. The RDSC provides standard datasets and conducts an output and publication control before any results can leave the RDSC. According to German law, formally anonymised data that is accessed in a secure environment satisfies the criteria of factual anonymization. Therefore, all external researchers who work inside the secure environment use factual anonymised data.
SUFs	The Bundesbank provides external scientific researchers access to factually anonymised micro data from surveys on households. The data is shared in the form of encrypted Scientific Use Files (SUFs) and can be downloaded by external researchers after the RDSC has approved the project request and shared a link as well as the password to unencrypt the files. For SUFs, the Bundesbank does not conduct an output or publication control.
B-Data	The Bundesbank provides external scientific researchers access to formally anonymised supervisory micro data on firms, banks and other financial institutions. The data can be accessed in the secure environment of the Bundesbank's RDSC only. Any access is subject to application and approval by the RDSC. The RDSC provides customised datasets and conducts an output and publication control before any results can leave the RDSC. According to German law, formally anonymised data that is accessed in a secure environment satisfies the criteria of factual anonymization. Therefore, all external researchers who work inside the secure environment use factual anonymised data.
PUFs	The Bundesbank provides everyone access to aggregated statistics in form of public use files (PUFs) on their website. For PUFs, the Bundesbank does require an application or approval by the RDSC nor does the RDSC conduct an output or publication control. Data that can be downloaded from the Bundesbank's website are absolutely anonymised. This means that they cannot be identified directly through labelling of reporting agents or other legal or natural persons, legal entities or branches (e.g. by means of their name or address or an officially issued identification code) or indirectly through deduction.

- **ES** - Banco de España

Two access modes have been identified in Banco de España: CAER (Controlled Access for External Researchers) and ES-EFF-PUF (Survey Household Finance Public Files).

The first one allows access to researchers after a Research Technical Evaluation Committee determine the applicant's suitability. The current datasets permit secure on site access, from Madrid and Barcelona Datarooms, and remote access to the researchers. In both cases, the extraction of information is only possible upon request and will be delivered after a thorough review to ensure that no personal and/or confidential information is disseminated. In addition, work is ongoing on the implementation of automatic output control tools.

So far the access to the second one, EFF public files, is provided to researchers through direct download from the EFF website after registration and acceptance of some terms of use. Public files are provided without identifying personal information.

- **EU** - European Central Bank

Darwin and ASTRA are document management tools that allow to make data available to users offering them the possibility of working collaboratively on documents and also downloading the data

- EU – Eurostat

Full details about available datasets and applicable procedures can be found on Eurostat dedicated website: https://ec.europa.eu/eurostat/web/microdata. Eurostat's PUF was not included in the answers at the time of the survey.

- MX – Banco de Mexico

Banco de Mexico has two Access mode: LISF (Financial System Information Lab) and EconLab.

LISF: Launched in 2020, its data access considers two stages. In the first one, access is allowed to anonymized microdata of the financial system to internal and external researchers in a coauthored scheme. Due to the pandemic, the access to data through the LISF is currently restricted to internal researchers. At the LISF, it is not possible to extract data and only aggregated results can be extracted, which are reviewed beforehand to avoid the identification and dissemination of sensible data.

EconLab: Provides external and internal researchers access to formally anonymized micro data. The data can be accessed in the secure environment on Banco de México and also with a secure remote access. The EconLab provides standard datasets and has an output and publication control of any results before it can leave the EconLab.

- **PT** – BPLIM - Banco de Portugal

Short name	Description
ISA (Internal Server Access)	Data can be provided to external researchers through an Internal Server. Access to BPLIM data using Internal Server implies the submission of a project. After approval, an account is opened in the server and the requested data copied to the account. The level of anonymization depends on the level of confidentiality of the dataset. Data can only be accessed in the secure environment of Banco de Portugal facilities. It is not possible to exchange files between the Internal Server and a local computer. Outputs must be requested and are all verified by BPLIM staff (or by internal researchers in case of co-authorship) before being sent.
ESA (External Server Access)	Data can be provided to external researchers through an External Server. Access to BPLIM data using External Server implies the submission of a project. After approval, an account is opened in the server and the requested data copied to the account. The level of anonymization depends on the level of confidentiality of the dataset. Data can be accessed remotely using a safe connection (the *nomachine* software) and it is not possible to exchange files between the External Server and the local computer. Outputs must be requested and are all verified by BPLIM staff before being sent.
SA (Surrogate Access)	External researchers not involved in joint projects with Banco de Portugal internal researchers can still gain access to datasets considered to be more sensitive (only available to internal researchers). To gain this access they need to submit a detailed project that will be evaluated according to the relevance of the topic for the research agenda of Banco de Portugal and must be approved by the Head of the Economics and Research Department. If accepted, a data expert at BPLIM will interact with the researcher and run the scripts (written by the researcher) on the original data. The data expert at BPLIM will use the resources available at the Internal Server (ISA). However, the researchers do not have access to the data (only to a randomized sample to help coding).