

## **Transnational Access to Confidential Microdata.**

Roxane Silberman and Maria Alkhoury (Centre d'accès sécurisé aux données / Secure Data Hub)

[roxane.silberman@casd.eu](mailto:roxane.silberman@casd.eu); [maria.alkhoury@casd.eu](mailto:maria.alkhoury@casd.eu)

### ***Abstract***

With legal frameworks changing, secure access to confidential microdata has grown significantly and highly detailed administrative and surveys data can increasingly be used for research and evidence-based policies. Among other modes (on site access, remote execution, synthetic data), remote access is favoured by the researchers who can directly manipulate the data. Facilitating access at national level, remote access also makes easier transnational access. While barriers remain and many countries are still hesitating, others have now more than 10 years experience providing remote access across borders.

In this paper, we will:

- 1) present a short overview of transnational access to confidential microdata with a focus on remote access mode: which countries, what data, to whom;
- 2) analyse how barriers have been overcome in France;
- 3) describe the research impact based on the CASD – the French Secure Data Hub – experience: data sources most used, types of projects, how they support innovative and comparative work, researchers' mobility, collaboration between research institutions from several countries; and
- 4) discuss the remaining difficulties: extension to other countries or other data, joint use of data and the prospects that are emerging based on the first lessons from the IDAN (international Data Access Network) project, a cooperative network of 6 Research Data Centres from 4 countries (France, Germany, Netherlands and the UK) that aims at building reciprocal access points between partners for facilitating the use of secured data from several countries for research.

# Transnational Access to Confidential Microdata: Progress and Impact for Research

*Draft version October 28, 2021 Do not quote without permission*

Roxane Silberman\* and Maria Alkhoury\*\*

\* Centre d'Accès Sécurisé aux Données (CASD & CNRS), roxane.silberman@casd.eu

\*\* Centre d'Accès Sécurisé aux Données (CASD), maria.alkhoury@casd.eu

**Abstract:** With legal frameworks changing, secure access to confidential microdata has developed significantly and highly detailed administrative and survey data can increasingly be used for research and evidence-based policies. Among other modes of access (onsite access, remote execution, synthetic data), remote access is favoured by researchers because it enables them to manipulate the data directly. Besides facilitating access at a national level, remote access also makes it easier to access data across borders. While barriers remain and many countries remain hesitant to adopt it, others now have more than 10 years of experience in providing remote access across borders.

In this paper, we will:

- present a short overview of transnational access to confidential microdata with a focus on the remote access mode: which countries, what data, to whom;
- analyse how barriers have been overcome in France;
- describe the research impact based on the CASD – the French Secure Data Hub – experience: data sources most used, types of projects, how they support innovative and comparative work, researchers' mobility, collaboration between research institutions from several countries;
- discuss the remaining difficulties: generalizability to other countries or other data, joint use of data and the prospects emerging based on the first lessons learnt from the IDAN (International Data Access Network) project, a cooperative network of 6 Research Data Centres from 4 countries (France, Germany, the Netherlands and the UK) which aims to build reciprocal access points between partners to facilitate the use of secure data from several countries for research purposes.

## 1 Introduction

Access to detailed and confidential microdata requires highly secure access conditions, which have been facilitated by the establishment of national Research Data Centres (RDCs). Only a small part of these microdata is gathered and harmonized at an international level, mostly at EU level by Eurostat while research increasingly focuses on using the whole range of government data sources from various countries, among these, administrative sources. Although access to these sensitive data across borders is developing, obstacles remain. Moreover, using confidential microdata from more than one source is a huge challenge, if they are fragmented in 'silos' between RDCs, which is particularly the case on an international level. All the above are major obstacles for comparative international research as well as for the development of evidence-based policy for the benefit of the public.

Access across borders faces different types of obstacles: 1) legal issues, as national laws might be "silent" regarding access across borders; 2) practical issues, as project

members may have to travel if no remote access is in place or if remote access across borders is not allowed; investment in translation into English or other languages may also raise difficulties; 3) objections which impede efforts to remove legal and practical obstacles: how much national data can be used by foreigners without the help of natives? To what extent can these data be useful, considering their comparability? Is there significant demand? For what kinds of projects?

Feedback from existing experience is therefore important in order to be able to move forward. In this paper, we analyse the experience of CASD, the French Secure Data Centre which enables remote access to a large set of confidential data across borders. CASD also participates in the International Data Access Network (IDAN) project, a network of 6 Research Data Centres from 4 countries (France, Germany, the Netherlands and the United Kingdom) which have been working together for 3 years to facilitate transnational remote access to confidential microdata.

This paper first situates CASD case within the current landscape and its recent developments regarding access across borders to secured data, before then providing an overview of the access from foreign institutions to the data available at CASD, showing how it accompanies international research cooperation. It qualitatively examines the various uses of the data in projects involving access and cooperation between institutions from different countries and finally underlines the investments and further developments required, so these possibilities can be fully exploited by users.

## **2 Access across borders to secured data: the current landscape**

To provide context for this French example, we begin by briefly describing the current landscape regarding access across borders to secured data for research use, with a focus on European Union developments (upon which France obviously depends as a Member State). This landscape is conditioned by the legal systems and interpretations thereof, and by the access systems in place.

Changes required for enabling access to confidential data for research were initially explored within the context of national legal frameworks. Since these were initially designed according to national needs, these legal frameworks, particularly statistical laws, have generally not covered the subject of access across borders. However, requests for access from researchers of foreign institutions began to emerge rather quickly, raising increasingly difficult issues as they had not been taken into account in initial national processes. One major question concerns how to sue and sanction a user beyond national borders in the event of a breach of confidentiality. This question was central to the lengthy discussions held by the Working Group on Statistical Confidentiality (WGSC) within the European Statistical System (ESS), which eventually supported the drafting of the 2013 European regulation on access to European microdata (based on the data of Member States). This led Eurostat to compare sanctions applied at national levels in the event of a confidentiality breach across the legal frameworks of the different

Member States in order to discuss equivalences. Formalizing equivalences and trust in contracts was also proposed as a solution in the 2015 OECD report by its Expert Group for International Collaboration on Microdata Access<sup>1</sup>. Within the European Union, recent developments associated with the General Data Protection Regulation (GDPR) are aligned with these discussions: they have facilitated transborder access within the EU yet require an adequacy decision for transferring data to non-EU countries. Such an adequacy decision has recently been made in the case of the United Kingdom following its roll out of Brexit. Nevertheless, on-going reservations and questions raised by several RDCs and National Statistical Institutes (NSIs) remain prevalent regarding how sanctions could be applied across borders. Concerning access requests from the United States, there is also uncertainty about the consequences of invalidation of the Safe Harbor agreement, then the Privacy Shield agreement by the European Court of Justice.

In addition to legal aspects, the issue of access across borders is also dependent on the modes of access and procedures implemented to ensure the data is processed under the security conditions required to ensure they remain confidential: on-site access, remote execution or remote access.

For confidential microdata, on-site access, which was historically the first mode of access, still remains the most broadly implemented mode today. On-site access requires researchers to travel, for more or less lengthy periods of time, often repeatedly. This raises both financial and compatibility issues with the organization of researchers' teaching loads. Non-resident researchers evidently face further difficulties. Nevertheless, on-site access remains the most commonly chosen mode of access when institutions begin to open up access to confidential data (INE in Portugal, Central Banks). Thus (for now at least) the number of on-site access facilities continues to increase. Over time, to address these mobility issues, on-site access has in several cases been decentralized within the country, with a network of onsite access centres implemented in the local offices of the NSIs (ISTAT in Italy with ADELE, DESTATIS in Germany, INE in Portugal) or in accredited centres in universities (StatCan in Canada, the Census Bureau in the United States). Nevertheless, even when access is decentralized, travel remains necessary for non-resident researchers. Moreover, authorization for foreign researchers to access data may entail specific extra requirements, such as membership in a local research institution. In the United States, access to Census Bureau data even requires minimum residence conditions in the United States (2 years).

Remote execution, which allows distant access via internet, raises fewer problems for cross-border access as no direct access to the data is required. It involves batch processing, whereby researchers can only perform set-ups without direct access to the data. The Luxembourg Income Study (LIS), providing access to income and wealth data from around 50 countries, has been implementing remote execution for many years and

---

<sup>1</sup> Chapter 4 *Establishing trusted partners in delivering microdata service* by Maurice Brandt and Chapter 5 *Sanctions for breach of confidentiality* by Brain Negin, Paul Jackson and Aleksandra Bujnowska

intends to continue to do so. Statistics Norway and the Australian Bureau of Statistics (ABS) are also strongly in favour of remote execution. However, this method has significant drawbacks and is not appreciated by researchers. Indeed, they are greatly constrained by the fact of not being able to directly explore the data and obtain intermediary outputs (not for publication), which would not respect the rules of anonymization. The outputs control of all intermediary outputs also engenders a heavy workload, which comes against the limits of the automatic control. Therefore, although this means of access facilitates access across borders, since direct access is only enabled for anonymized results, these systems are not well suited for complex statistical processing nor for joint analysis of data sources held in various national RDCs. This mode of access does not seem to be developing strongly. StatCan and IAB both use it as a complementary mode in combination with direct onsite access to secured data. Eurostat considered it as an option and then dropped it in favour of remote access.

Remote access, a true online mode enabling interactions with the system, enables researchers to "see the data" without being able to download it, working inside secure bubbles where codes and intermediate outputs can be stored until the analyses are finalised. Only final outputs intended for publication must meet the anonymization criteria. This mode, developed by pioneer centres as Statistics Denmark in 2001, adopted by CBS in the Netherlands in 2006 and CASD in France in 2010 (after an experimental phase between 2008-2009), is liked by researchers, as is evidenced by discussions held within the research community during the DwB project (Data without Boundaries, 2011-2015)<sup>2</sup>. More RDCs have adopted this mode of access for resident researchers, including the UKDS and, more recently, Finland Statistics with its FIONA system. IAB in Germany implements remote access alongside its on-site access and remote execution system, Joshua. The ONS-SRS facility, initially accessible through the UK IT government network, has recently begun to implement remote access open to UK universities. Other institutes, such as Statistics Austria, are considering moving to such a mode of access and the creation of a remote access to the Secure Use Files for European microdata is expected soon, after many years of discussions within the ESS since the adoption of the 2013 regulation which made it possible.

Nevertheless, remote access does not automatically entail access across borders. Decisions and procedures vary. CBS in the Netherlands offers remote access for researchers both within and beyond EU territory. CASD was recently allowed to open up access beyond the EU to North America with some specific conditions. Statistics Denmark requires those accessing data to be affiliated with a home research institution. Finland currently relies on the Eurostat list of accredited institutions. IAB in Germany is more restrictive. Remote access from other countries currently does not offer access of the same level as that available in Germany. However, within the context of the IDAN project, an agreement was signed with CASD in France to allow data to be accessed with the same level of detail from an access point implemented at CASD. Regarding the

---

<sup>2</sup> *DwB WP 4* <https://web.archive.org/web/20180830181742/http://dwbproject.org/about/>

case of the UK, remote access from foreign countries to the secured data available at ONS-SRS and UKDS<sup>3</sup>, currently not possible, is being debated. Still within the context of the IDAN project, remote access from IAB has been established as a first step. Interpretation of remote access plays - and will continue to play - a key role in differences and changes. While some consider that remote access is not a transfer of data, as users only view the data, which remain in the RDC systems, others consider that even if no physical transfer occurs, remote access should be considered as a transfer of data in terms of security. Within and across the EU, the GDPR has now confirmed this interpretation that may require adjustments for RDCs offering remote access from non-EU countries. An adequacy decision or specific contracts might then be necessary.

Several projects have also emerged to facilitate the use of secured data from multiple countries. The NordMAN project, which ended in 2017, involved several Nordic countries, which all agreed to improve access to cross-Nordic data facilitating thus transfer of data between them. As a first step in the IDAN project, 6 RDCs (CASD in France, IAB and GESIS in Germany, CBS in Netherlands, and ONS-SRS and UKDS in the UK) are developing agreements for implementing reciprocal provision of access points, enabling researchers to remotely access secured data provided by partner countries from their local RDC, all access points being installed in the same physical location.

This paper does not discuss the developing work on synthetic data, which aims to solve the problem of data anonymization not by deleting information, but by disturbing the data, an alternative to secure access systems that would also offer a wide access across borders. Requiring a lot of work and adaptations depending on the objectives of the analysis, synthetic data still raise difficulties as results may differ from those obtained with undisturbed data and remain difficult to implement widely.

As can be seen, questions and difficulties remain unsolved regarding access across borders to secured data. In many cases, it still necessitates travel, which has been (and still is) severely restricted by the pandemic and could remain problematic in the future with regards to decarbonization objectives. Experience from RDCs providing cross-border remote access should help to address this issue. The remainder of this paper focuses on one of these: the example of CASD in France.

### **3 Cross-border access to secured French data: CASD as an example**

Set up in 2010, by the National Statistical Institute, INSEE, CASD (Centre d'Accès Sécurisé aux Données), the [Secure Data Hub](#) has, since 2018, been a public institution bringing together the State, represented by INSEE (French NSI), and several research institutions and universities: GENES, CNRS, École Polytechnique and HEC Paris. The purpose of CASD is to organize and implement secure access services to confidential data

---

<sup>3</sup> Currently for one data source, “*Understanding Society*”

for non-profit research, study, evaluation or innovation activities. Acting as a third party between data producers and users, CASD offers a wide range of data covering diverse topics, from public and private data producers. Data from the public sector, surveys and administrative data are mainly sourced from the French National Institute of Statistics and Economic Studies (Institut National de la Statistique et des Études Économiques – INSEE), the statistical departments in ministries, the Public Finances General Directorate (Direction Générale des Finances Publiques – DGFIP) affiliated to the Ministry of Finance and Public Accounts which holds important tax data, plus other public and private institutions. CASD also hosts health data including those pertaining to epidemiological cohorts supervised by academics. The potential to associate selected survey and administrative data sources with names as well as with National Registration Numbers/Social Security Numbers (Numéro d’inscription au répertoire/Numéro de sécurité sociale – NIR) also exist.

### **3.1 CASD remote access modalities**

Set up relatively late compared to many other countries in which onsite access has been implemented since the 1990s and remote access developed from the beginning of the 2000s, CASD benefited from the technological advances made elsewhere, opting without delay for remote access, enabling researchers to directly manipulate the data. The solution designed and fully controlled by CASD is an integrated system, easy to deploy and accessible only through a dedicated secured access terminal (the SD-Box access point) and biometric authentication. The SD-Box is an autonomous terminal, the sole purpose of which is to allow access to the highly secured infrastructure where the data are stored (the “bubble”). Data cannot be downloaded from the “bubble”, but can be processed in virtual research workspaces dedicated to each project. More than one user and several projects may share the SD-Box, each virtual environment being completely isolated from others. The high level of security is combined with great ease of work for the user until the final drafting of the articles inside the secure bubble of the project. In fact, the secure bubble includes a wide range of scientific and data analysis software tools as well as enough space and computing power to match the users’ needs. Final outputs are controlled by CASD according to the data producers’ rules for anonymization.

Access to the data is subject to authorization along the main lines of the Five Safes<sup>4</sup>. CASD is not in charge of accreditation. Indeed, an independent authority, including representatives of the data producers, the Statistical Confidentiality Committee (*Comité du secret statistique*, CSS), ensures projects and researchers are “safe” in coordination with the bodies in charge of data privacy protection (Commission Nationale Informatique et Libertés, CNIL and National Archives). Once project members are approved, if they need to work directly on the data, they are obliged to attend a CASD

---

<sup>4</sup> The Five Safes framework includes five dimensions: safe researcher, safe project, safe data, safe settings, safe outputs. See Desai, Tanvi, Felix Ritchie, and Richard Welpton. “Five Safes: Designing Data Access for Research.” Working Paper. University of the West of England, 2016. (<http://eprints.uwe.ac.uk/28124>).

training session on security and confidentiality aspects (the attendance to this session being valid for 4 years). After this session, they sign CASD terms and conditions of use (which set out the rights and obligations of the signatory when using CASD services) and receive an access smartcard on which their fingerprints are encrypted. Upon completion of this process, they are considered to be “enrolled” and they become CASD users.

CASD is in charge of the contracts. CASD contracts govern 2 different aspects: the conditions for hosting SD-Box access points and aspects relating to the services provided to research projects. Each aspect entails the payment of fees which can be covered either by the same institution or by different institutions. Thus, institutions can be differently involved as a host for a SD-Box and as a participant in a project. Similarly, a user, once enrolled can use the SD-Box of their institution or an SD-Box from another institution involved in the project and hosting an SD-Box or any other institution they may be visiting where an access point is available. Once projects are closed, the SD-Box is sent back to CASD if no other project is using it. However, an increasing number of institutions are tending to keep and pay for the SD-Box access points even if there is no ongoing project, as they consider it to be an asset for potential new projects. Indicators on access set out hereafter within this paper evidently depend on this system and its procedures.

The total number of SD-Box access points deployed has dramatically increased over time, starting with a handful in 2010 to **730 SD-Boxes** currently hosted by institutions all over the globe.

### **3.2 Opening access across borders**

CASD operates within the legal framework of France (mainly the 1951 statistical law and 1978 *Informatique et Libertés* law for the protection of personal law). Both were revised several times, in particular to allow derogation for access to confidential data for research, statistical and historical purposes. As in many countries, both were “silent” regarding access across borders. Facing an increasing demand from researchers, in 2011 INSEE allowed CASD to provide secure remote access to their confidential data from EU countries. The same conditions and procedures for authorization (approval by the CSS and enrolment at CASD) as for resident researchers in France apply. This decision was based on the high security provided by CASD, and on an analysis of the conditions of international assistance for criminal prosecutions in case of a breach of confidentiality that would have repercussions in France.

In February 2018, INSEE as well as the office of Statistics and Prospective Analysis – SSP (Service de la Statistique et de la prospective) affiliated to the ministry of Agriculture – opened access to their data to researchers in the United States and Canada. The same conditions apply to researchers who are citizens of a Member State of the European Union when working at a North American university or research centre, while



other researchers working at a North American university need to be part of a research project in partnership with a research centre or university in a Member State of the European Union. Since this decision, several other statistical departments in ministries have also authorized access to their data with SD-Box access points located in the United States and Canada under the same conditions.

As indicated above, CASD is also a member of the IDAN project. A specific agreement between CASD and the ICPSR at Michigan (US) also aims to facilitate access to French data from the US for researchers working on other confidential data available at ICPSR.

The adoption and entry into force of the GDPR throughout the EU has naturally impacted conditions of access to the French data. As set out above, the GDPR consolidated access from the EU and AELE countries, setting up a common legal framework for the protection of personal data which facilitates transfer or personal data within the EU, while requiring an adequacy decision (in the case of the UK, which was adopted in June 2021) or specific contracts.

### 3.3 A first approach: the location of CASD SD-Box access points

Within this context, many SD-Box access points are now located in various countries. Since 2017<sup>5</sup>, 217 access points have been requested to be deployed abroad, representing 18% of the 1220 requests for SD-Boxes over the period. The percentage is rather similar (16%), if we just consider the access points still active today, i.e. the boxes currently hosted by institutions (120 of a total of 730).

Year	2017	2018	2019	2020	2021	Total <sup>1</sup>
Number of SD-Boxes requested abroad	9	55	59	50	44	217
Number of SD-Boxes requested in France	147	289	184	198	185	1003

Table 1: Number of SD-Boxes requested abroad and in France according the year in which they were requested

1: The total does not represent the number of physical SD-Boxes but rather the flow of their hosting: one SD-Box could be hosted multiple times

Table 1 shows the regularity of the number of boxes requested abroad each year over the same period between 2018<sup>6</sup> and 2021, with a slight decline since 2020 most probably due to the Covid-19 pandemic situation.

Country	Total (and active) SD-Boxes	Percentage
United Kingdom <sup>7</sup>	45 (16)	20.6%

<sup>5</sup> In 2017, the SD-Box management process was dissociated from that of secure access services provided to research projects: this is why SD-Box numbers before 2017 are unavailable.

<sup>6</sup> Due to a change in CASD procedures in the middle of 2017, the number of boxes is underestimated for this year.

<sup>7</sup> In the case of the institutions in the UK, the great decrease in the number of SD-Boxes might reflect in part an anticipation of the impact of Brexit, as it was uncertain whether remote access would remain possible.

Germany	34 (18)	15.6%
Italy	30 (19)	13.8%
Switzerland	28 (14)	12.8%
United States	21 (11)	9.6%
Netherlands	11 (8)	5.1%
Belgium	10 (9)	4.6%
Spain	10 (4)	4.6%
Sweden	9 (6)	4.1%
Canada	4 (3)	1.8%
Denmark	4 (4)	1.8%
Austria	3 (2)	1.4%
Ireland	3 (2)	1.4%
Luxembourg	2 (1)	0.9%
Finland	1 (1)	0.5%
Norway	1 (1)	0.5%
Portugal	1 (1)	0.5%
<b>Total</b>	<b>217 (120)</b>	<b>100%</b>

Table 2: Foreign countries that have hosted (or currently host) SD-Boxes

Though a limited number of countries host about 70% of the SD-Boxes that have been sent abroad over the period (see Table 2), mainly the UK, Germany and Italy, the number of countries hosting SD-Boxes increases year on year. Indeed, each year, CASD receives requests to host SD-Boxes in a new European country: for example, Luxembourg and Norway in 2019, Finland in 2020 and Portugal in 2021. The opening of access from North America in 2019 immediately resulted in SD-Box hosting requests (12 were sent in 2019).

The number of research institutions hosting SD-Boxes in these countries also increases regularly and a couple of these host a large number of access points (see Table 3).

Institution	Country	Number of total (and active) SD-Boxes	Percentage
London School of Economics	United Kingdom	18 (5)	8.3%
Bocconi University	Italy	16 (11)	7.4%
University of Warwick	United Kingdom	7 (0)	3.2%
Stockholm University	Sweden	6 (3)	2.8%

University of Mannheim	Germany	6 (4)	2.8%
University of St. Gallen	Switzerland	6 (1)	2.8%
University of Zurich	Switzerland	6 (4)	2.8%
ETH Zurich	Switzerland	5 (3)	2.3%
University of Louvain	Belgium	5 (5)	2.3%
Erasmus University Rotterdam	Netherlands	4 (2)	1.8%
Other institutions	Other countries	138 (82)	63.6%
<b>Total</b>		<b>217 (120)</b>	<b>100%</b>

Table 3: Top 10 Foreign institutions that have hosted (or currently host) SD-Boxes

### 3.4 A second approach: foreign institutions involved in projects

However, the location of the SD-Box does not fully reflect the use of the French data by researchers from foreign institutions. Foreign institutions can be involved in various ways in projects using the data available at CASD. They can be the requesting institution<sup>8</sup> for the project. They can participate in a project with other institutions from other countries, including France. They can be involved in projects while not hosting an SD-Box as the researchers could use access points located in a partner institution in France or in another country. A broader overview of the involvement of foreign institutions in the projects is therefore an important aspect when discussing the role of secure remote access across borders within the context of the development of international cooperation in research.

In 15.5 % (164 projects) of all CASD projects (1055 projects), the requesting institution for the project is located outside of France. Note that in most cases, even if many institutions are involved in the project, only one institution will be the main requesting one in the French accreditation process. Nevertheless, 18 projects have requesting institutions located in 2 or 3 different countries, reflecting the ongoing expansion of international research

Table 4 shows the increase in the number of foreign institutions applying as the requesting institution in the project application (from 4 projects between 2010 and 2012 to 25 projects in 2020).

Location of the requesting institution(s)	Application year										
	2012 and before	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total
<b>Abroad</b>	4	9	11	7	16	19	24	25	26	23	<b>164</b>

<sup>8</sup> The requesting institution is the institution of the application procedure. This information is collected by the CSS. It is the institution which is liable when it comes to processing of personal data.

<b>France</b>	123	79	46	84	81	80	89	96	127	86	<b>891</b>
<b>Total</b>	127	88	57	91	97	99	113	121	153	109	<b>1055</b>

Table 4: Projects according to the location of their requesting institutions and their application year

Independently of the role of the institution in the application and of the location of the SD-Box access points, and considering all project members (whether directly working on the data and enrolled or not), 11% (469) of the project members in a CASD project of a total of 4345 are affiliated to a foreign institution.

The involvement of these researchers in projects held at CASD is important as can be seen in Table 5 showing different cases in international cooperation: projects only involving researchers affiliated to one foreign country, projects involving researchers belonging to two or more countries including France or a different country.

<b>Number of countries</b>	<b>Number of projects</b>	<b>Percentage</b>
One country: France	811	76.9%
One foreign country	100	9.5%
Two countries	119	11.3%
Three countries	21	2.0%
Four countries	4	0.4%
<b>Total</b>	1055	100.0%

Table 5: Projects according to the number of countries of the institutions of their members

The table does not reflect the place of access – as a user may access CASD data from any SD-Box – but it sheds some light on the overall international cooperation involved in the projects, which has been in part supported by the development of cross-border access, as will be shown in the next section based on a more qualitative analysis.

### **3.5 What usage of the French data in projects involving foreign institutions?**

Since access from abroad to data available at CASD was made possible, we have seen an increase in the number of SD-Boxes installed in foreign countries, which reflects the increasing involvement of foreign institutions in the projects with various configurations (alone, or with France, or with other countries). For what usages? The answers are important for advocating the development of remote access across borders to national secured microdata, which are not harmonized on the international level (particularly those not collected by Eurostat). Drawing on CASD experience, we can observe three main usages.

### **3.5.1 Supporting researchers' mobility**

As detailed above, once a researcher has been enrolled by CASD, becoming thus a user, and is in possession of their access card for authentication, they can use any SD-Box. Therefore, if a user is visiting another institution holding an SD-Box, they can use it if agreed by this institution. As the SD-Box access points are now installed in various institutions regularly involved in projects at CASD, if a user wishes to use an SD-Box in a different institution, they only need to contact CASD if no SD-Box is available or if the institution does not authorise their access. An analysis of the emails received by CASD in such cases provides an understanding of CASD users' mobility and how access from abroad is crucial to support this. We can broadly distinguish 3 main cases:

1) short visits for meetings, workshops or conferences (linked or not linked to projects). In this case, researchers need to access the data for discussions with colleagues or for finalizing a paper. These cases are likely underestimated: visitors in institutions involved in their projects will frequently find an SD-Box and therefore not contact CASD, or for very short visit, users might not bother contacting CASD if no SD-Box is available;

2) users moving (mainly from France but also from other countries) for a PhD or post-doc position in an institution abroad: they are either still working on a project based on French data or they may be starting a new project;

3) users moving abroad for a teaching position and still working on or intending to work on French data.

Quite clearly, for PhD, post-docs and teaching positions (about 60% of the reasons for contacting CASD in case they need an access point in another institution), enabling remote access across borders to secured data proved to be crucial to support the increasing mobility of researchers internationally.

### **3.5.2 Use of original and rich data sources**

A second aspect of projects involving foreign institutions and cross-border access to the data, are projects on subjects which are international by nature, such as international trade, import and export, the globalisation of firms, geographical localisation of firms, international mobility. These mostly require data on firms, many of requiring secure access, which has been possible in France comparatively early and on a wider scale than in many other countries.

More broadly, business data, including employer-employee data are in the top ten types of data used by projects held at CASD on a great variety of topics, particularly on the labour markets. In addition to this, tax data made available for research in 2013 are also rich resources not frequently available in other countries and attracting research involving foreign institutions.

### **3.5.3 Comparative work**

It is not easy to obtain an exhaustive list of all CASD projects involved in comparative research, as researchers are not required, when applying for access to French data, to indicate whether it is linked or not to a project conducted in parallel in another country. However, some projects with a comparative approach are easily identifiable through their title. Focusing on these projects (10) we can see that they deal with diverse subjects, comparing different countries, with France being a country commonly used for comparison: for example one project compares housing construction between Paris and New York, another looks at working hours between France and Germany, a third analyses how wage distribution compares between the UK and France.

An interesting case to study is the collaboration between CASD and the IAB German Research centres in the framework of the IDAN network. In this collaboration, an access point for IAB data was deployed in CASD premises. Between August 2018 and September 2021<sup>9</sup>, CASD has welcomed 10 different users to work on IAB data. Seven out of these 10 users are also CASD users. Some began working on IAB data after having worked on CASD data, and some requested access to CASD data after having worked on IAB data and made contact with different CASD teams. By looking at the application of these projects alone, we are unable to clearly identify a comparative approach due to the fact that researchers are not required to specify this point in their data access application. However, all these projects work on employment and labor issues which is a common theme between the data these projects have access to: in fact, IAB provides access mainly to data in the field of social security and employment.

The 3 broad types of usages identified above are linked in many cases with researchers' mobility, facilitated by remote access facilities across borders which plays an important role in the development of use of rich French data sources for innovative research abroad and for comparative projects. The London School of Economics in the UK and Columbia and Princeton Universities in the United States, privileged destinations for PhD, post doc students and researchers, well illustrate the role of mobility in the development of projects involving these institutions.

### **3.6 The need for continuing investments**

What data sources can I use? What data sources could be compared for my work? Could I use data sources from the various countries featured in my analysis? These are three common questions researchers have in mind when looking for data from other countries.

Evidently, language is an initial obstacle and, while in many cases foreign researchers tend to work with native colleagues, translation in English are essential for supporting cross-border access. An English version of the various sections of the CASD website is

---

<sup>9</sup> Due to the Covid-19 situation, the CASD safe room was closed from March 2020 until May 2021. Three new different users requested to book the Safe room, but we were obliged to refuse these booking requests.

now available and enables researchers to search data sources in the catalogue by topic in English. Data source abstracts are increasingly translated into English.

Another obstacle in the case of CASD is the enrolment process. Researchers who reside in another country and wish to work on the data must travel to France for their biometric enrolment. Consequently, less researchers affiliated with foreign institutions enrol than accredited researchers affiliated with French institutions: around 59% of members of an active project affiliated with a foreign institution have an access card, compared to 70% when the affiliate institution is located in France. As a result, CASD is developing a solution to enrol remotely accredited researchers, in order to facilitate access for foreign researchers to CASD data.

Comparability is a third central issue regarding the use of national data sources when these are not harmonized within an international context, as is the case of European microdata. This first concerns information on existing data sources. A tentative tool for all EU countries, [CIMES](#) (Centralising and Integrating Metadata from European Statistics), developed by CASD in partnership with CNRS, has been developed within the DwB project, raising interest from Eurostat. With a smaller perimeter, the IDAN project provides links to partners' catalogues and a list of their top ten datasets used by researchers on its website. Furthermore, CASD has worked with IAB to increase and simplify the feasibility of cross-country comparative research between Germany and France. For this purpose, they published a [report](#) together in which they present two comparison tools. The first of these tools provides an overview of datasets available at the FDZ, with similar datasets available at the CASD according to their subject focus. The second tool compares two similar administrative datasets in more detail, providing an overview of their characteristics before comparing them on individual variables. Combined use of these data for comparative projects is facilitated by the reciprocal provision of access points between IDAN partners, enabling the researcher to access the data of all IDAN partners from the same physical location in their local premises.

## **4 Conclusion**

Cross border access to confidential microdata for research still today faces many obstacles and objections. The example of CASD shows how remote access across borders can be secured while also supporting the use of rich national data resources for innovative or comparative research, together with both the increasing mobility of researchers and the development of international cooperation in research.