6 The LMU-ifo Economics & Business Data Center (EBDC)

Sebastian Wichert, Heike Mittelmeier, Valentin Reich, Michael Rindler

6.1 General Information about the EBDC

The LMU-ifo Economics & Business Data Center (EBDC)[1] is the joint Research Data Center (RDC) of the ifo Institute, the LMU Munich's Department of Economics[2], and the LMU Munich School of Management[3]. It was founded in 2008 with the financial support of LMU's long-term strategy program "LMUexcellent". In 2011, the EBDC was accredited as an RDC by the German Data Forum (Rat für Sozial- und Wirtschaftsdaten (RatSWD))[4]. Therefore, the EBDC provides non-discriminatory access to its research data free of charge, is subject to annual monitoring, and participates in a standardized, impartial mechanism for resolving user complaints, among other things. In 2014, micro-data from official statistics became available at the EBDC through cooperation with the Bavarian State Office for Statistics. The EBDC is well connected in the research data sphere in Germany and beyond: It is i.a. part of the national research data infrastructure (NFDI), the Statistik Netzwerk Bayern as well as Leibniz research network "LeibnizData". In these networks, the EBDC and its partners are promoting better access to and linking of research data.[5]. In order to better address the growing importance of research data and the diverse tasks it has taken on since its foundation, the EBDC has been continuously expanded in terms of personnel and infrastructure and has become an independent research-oriented infrastructure and service center within the ifo Institute since 2022.[6]

When the EBDC was founded, its primary goal was "to establish a central place for the collection and supply of research data for economic and business sciences". More specifically, the main task was to make ifo's long-running surveys available to the broader research community since this data until then was only partially accessible within the institute. To this end, the EBDC team developed and maintains to this day well-documented, regularly updated, and pseudonymized research micro-data sets from ifo's surveys. After a retention period, these data sets are provided to ifo researchers and guest researchers only for non-commercial

[1] More, up-to-date information about the EBDC can be found on their webpage: `www.ifo.de/en/ebdc`.
[2] See `www.en.econ.lmu.de`.
[3] See `www.som.lmu.de/en`.
[4] See `www.konsortswd.de/en/ratswd`.
[5] See `www.konsortswd.de/en`.
[6] Earlier comments on EBDC can be found in Abberger et al. (2007), Becker and Wohlrabe (2008), Seiler (2012) and Mittelmeier (2020).

academic research projects. They can be accessed in a safe and well-equipped environment at EBDC to secure the confidentiality of the survey responses. In addition to the micro-data from ifo's surveys, the EBDC team has developed and continually enhances innovative linked research data sets: the EBDC Business Panels. These data sets contain survey data from both the ifo Institute and external balance sheets. While the provision and further development of the (linked) ifo data sets is still an important task of the EBDC, the EBDC's functions have expanded considerably since its early days. Today, the EBDC considers itself a research-oriented service and infrastructure provider. Its mission is to enable and participate in innovative and challenging empirical research projects, especially with large unconventional data and/or new methods. The EBDC Advisory Board provides strategic guidance to the ifo Executive Board and EBDC management to continuously improve the EBDC and its services.

As of early 2023, the EBDC has hosted the following regularly conducted ifo survey data sets, which are all described in previous chapters:[7]

- ifo Business Surveys (Manufacturing industry, Construction industry, Trade, Service Sector)
- ifo Business Survey in the insurance industry
- **EBDC Business Panels** (EBDC Business Expectation Panel, EBDC Business Investment Panel, EBDC Business Innovation Panel)
- World Economic Survey/Economic Expert Survey
- ifo HR Survey
- ifo Architects Survey
- ifo Education Survey
- FamData – Database and Survey

The following sections of this chapter provide an overview of services and data that the EBDC offers to its external guest researchers. The EBDC Business Panels are presented in detail, followed by data protection measures during the preparation of the data sets and during use by guest users, access regulations, the EBDC research data repository, and lastly the branch office of the Research Data Center of the Statistical Offices of the Länder at the EBDC.[8]

[7] Further, irregularly conducted or discontinued surveys and data sets that are processed and hosted by the EBDC are the ifo Investment Surveys (Manufacturing Industry, Trade, Leasing), the ifo Innovation Survey, the ifo Investment Database, and several meta-surveys. Further original and/or replication data sets are available through the EBDC repository (see Section 6.6) or the ifo webpage (e.g. iPEHD – the ifo Prussian Economic History Database).

[8] Other ifo survey data sets such as the ifo Business Survey or the ifo Education Survey, which are provided by the EBDC, are explained in detail in other chapters of this handbook. Therefore, they are intentionally omitted here. Furthermore, the EBDC provides further services and fulfills other tasks in the areas of research data infrastructure management, Open Science, networking and committee work, and further education. More, up-to-date information about these services, all data sets available at the EBDC and their access regulations as well as all publications with them can be found on the EBDC's webpage: `www.ifo.de/en/ebdc`.

6.2 The EBDC Business Panels

Overview

In the EBDC Business Panels, namely the EBDC Business Expectation Panel (BEP), the EBDC Business Innovation Panel (BINP), and the EBDC Business Investment Panel (BIP), each micro-data set from the ifo surveys, namely the ifo Business Survey (IBS), the ifo Investment Survey (IVS), and the ifo Innovation Survey (INS) - is combined with standardized financial and structural data from the very large external company database Orbis provided by Bureau van Dijk through a machine learning based record linkage procedure. As of 2023, the Orbis database contains more than 450 million companies worldwide, of which 45 million have detailed financial information with hundreds of variables from 170 local data providers.[9] For Germany, the data is delivered by Creditreform and contains i.a. information of the German Commercial Register ("Handelsregister"), the Federal Gazette ("Bundesanzeiger") as well as balance sheet data and win and loss statements from company reports and information from the press. For the EBDC Business Panels more than 50 positions were selected or recalculated from these sources. Through the record linkage, subjective company-specific expectations, estimates, and plans from the ifo survey data can be related to the objective, realized balance sheet and structural data in the company database. The EBDC Business Panels are published yearly with a one-year retention period. Table 6.1 gives an overview of the EBDC Business Panels and the underlying ifo surveys.

Structure

The structure of all EBDC Business Panels is very similar. The EBDC Business Panels contain the same three sets of variables as the respective ifo surveys: 1) identification variables (e.g. pseudonymous ID, survey month, etc.), 2) variables from regularly asked questions (e.g. business expectations), and variables from special irregularly asked questions (e.g. questions on effects of the Covid-19 pandemic or the war in Ukraine on business performance). If a company produces several goods or provides several services, it may fill in two or more questionnaires for different products/services or product/service groups. This is only the case for the ifo Business Survey. Therefore, it is possible that several survey observations of a company for different products or product groups are assigned the same financial information from Orbis. Since the ifo Investment Survey does not differentiate between products, each ifo survey participant is assigned to exactly one firm from the Orbis data base. Of course, this only applies if the firm is contained in the Orbis data base and can be linked with the record linkage procedure. Identification of an observation in the EBDC Business Panels is

[9] See `https://www.bvdinfo.com/en-gb/our-products/data/international/orbis` for more information. Due to contractual obligations, the Orbis database is only available for ifo and LMU researchers. In earlier years the Amadeus database by Bureau van Dijk and/or the Hoppenstedt database (today: Dun & Bradstreet Firmendatenbank) have been used as well. For each company, the balance sheet information of the provider with the most balance sheet years is preferred.

Table 6.1: Overview of ifo surveys and EBDC Panels

Survey	Period	Observations	Companies	thereof in EBDC panel
Business surveys:				
Industry	since 1980	1,775,268	18,946	7,445
Construction	since 1991	1,102,681	5,958	2,693
Trade	since 1990	873,718	11,530	5,299
Service provider	since 2004	485,140	9,378	6,578
Insurance	since 1999	46,213	156	-
Investment survey:				
Industry	since 1964	393,528	23,901	7,076
Trade	2000 - 2014	15,696	4,546	-
Innovation survey:				
Industry	1982 - 2015	41,996	7,862	3,060

As of early 2023.

based on a company EBDC ID, the observation year and other data set-specific time variables. The financial information in the EBDC company panels is based on individual rather than consolidated financial statements when available.[10] However, the financial statement data from the company data bases has not simply been adopted. Rather, a new EBDC balance sheet scheme was developed which integrates variables from all company data bases and abstracts them from the existing differences in the original data sets.[11] The EBDC balance sheet format is based on the balance sheet and income statement structure of the German Commercial Code (HGB) and in some cases also shows variables according to the total cost method or the cost-of-sales method.[12] For a detailed and appropriately structured overview of the available balance sheet and income statement variables, please refer to the respective variable list of the corresponding EBDC Business Panel.

EBDC Business Expectations Panel

For the EBDC Business Expectations Panel (BEP), the link results in a combination of monthly data from the ifo Business Survey and annual data from the financial company data bases. The BEP-ID consists out of three components: the variable “bep_company_id”, which is a serial company number, the variable “questionnaire_id”, which consecutively numbers the questionnaires per firm, and the variable “sector_id”, which contains the sectors surveyed. The

[10] The respective trade type is indicated by the variable “reporting_basis”. Limited financial data means here that the balance sheet information was not published, but mostly requested individually.
[11] The conversion scheme used to convert the original variables into the newly generated EBDC data balance sheet variables has been used, can be viewed at the EBDC. A detailed standard balance sheet is also available, which can be used for orientation purposes.
[12] Upon request and in exceptional cases, original balance sheet variables from Orbis can be provided.

ifo Business survey in the manufacturing industry (KTVG) asks for products, the ifo Business Survey in the service sector (KTDL) for service branches, the ifo Business Survey trade (KTHAN) for product groups and the ifo Business Survey in the construction industry (KTBAU) for construction types. One questionnaire can contain questions for several divisions.

The EBDC Business Expectation Panel data set is sorted by BEP-ID, year, and month and hence provided in a (unbalanced) long format. Each observation in the survey can be uniquely identified by these three variables, of course without disclosing the name of the firm. Since the ifo Business Survey is conducted monthly, up to 12 monthly reports per BEP-ID can be available for each year. This is followed by the balance sheet information in an additional month denoted "99" constructed for this purpose. This data set structure facilities the handling of monthly and annual information and keeps the size of the data set comparatively small. The data set contains the variables ordered by their function: 1) Identification variables, 2) balance sheet and P&L information from the company data bases, and 3) the ifo survey variables. In addition to the BEP-ID, the year and month, the identification variables also include information on industry codes, size classes of persons employed, federal state, stock exchange listing, legal form, etc.[13] The variable lists provided by the EBDC give an overview of all variables, their significance and special features, and contain all the questions of the individual ifo Business Survey, including the survey period and periodicity.

EBDC Business Investment Panel

Due to the semi-annual cycle of the ifo Investment Survey (IVS), the EBDC Business Investment Panel has a slightly different structure than the EBDC Business Expectations Panel. The data set no longer contains monthly data, but a variable "season" indicating whether the data come from the spring (1) or autumn (2) survey or from a balance sheet (99). The data set is sorted by BIP-ID ("bip_company_id"), "year", "season". The rest of the data set structure matches that of the EBDC Business Expectations Panel. The IVS was discontinued in 2022 and the last survey wave was conducted in spring 2022. However, the data sets of the IVS are available at the EBDC just like all other ifo research data sets.

EBDC Business Innovation Panel

The annual ifo Innovation Survey (INS) refers to individual products of a company like the ifo Business Survey (IBS) for the manufacturing sector. Participants for this survey were sampled from the IBS. Therefore, these two data sources together form the EBDC Business Innovation Panel (BIP). The INS surveyed both innovation activities and objectives as well as innovation impulses and obstacles. Product and process innovations were addressed, and there were changing special questions in specific years. The INS is attached to the IBS from the manufacturing sector as an annual data set with the month specification "98", just like the balance sheet data with the month specification "99". The INS was discontinued in 2017 and

[13] For reasons of anonymity, the federal state information was deleted for large companies ($>$ 10,000 employees).

the last survey wave was conducted in 2016. However, the data sets of the INS and the BIP are available at the EBDC just like all other ifo research data sets.

Record Linkage

To link the ifo surveys to balance sheet data from Orbis, the address information of firms in both data sources is accessed. The procedure ultimately leads to a mapping table that allows to combine information from the ifo surveys and Orbis into the EBDC Business Panels.

Since there is no common identifier between the two data sources, a probabilistic linkage procedure has to be applied based on various similarity metrics. To this end, the EBDC uses a five steps procedure:

(i) The records containing name and address information are cleaned and standardized.

(ii) A set of pairs that are considered potential matches is identified based on simple heuristics.

(iii) For each pair in this set, various similarity metrics are computed.

(iv) A supervised machine learning algorithm is trained on the similarity scores of a set of pairs that were manually labeled as matches or non-matches as training data. For the remaining pairs, this algorithm uses their similarity score to predict whether the pairs refer to the same entity.

(v) Predicted matches are manually reviewed and corrected if necessary.

The linkage is especially successful for companies that have been added to the survey since around 2005 with nearly perfect coverage for the most recent years. Figure 6.1 illustrates the process of creating the EBDC Business Expectations Panel. More information about this procedure and the results can be found in Reich (2023).

6.3 Data Protection and Data Security Measures during Processing and Usage at the EBDC

Legal data protection, technical data security as well as the confidentiality of the survey participants and their answers have the utmost importance for the researchers at the ifo Institute and the EBDC team.[14] This is reflected in the way how the research data sets are created from the raw survey data by the EBDC team as well as in strict data usage regulations

[14] All ifo researchers, and in particular the EBDC team, are dedicated and bound to confidentiality by their employment contract and the obligatory principles of good scientific practice.

for guest researchers.[15]

After completion of their surveys, the ifo researchers in charge transmit their data to the EBDC, usually in a standardized process (i.e. within a predetermined frequency, content, and granularity).[16] In the case of the ifo Business Survey, the EBDC team gets access to the raw micro-data (i.e. the survey answers), only after the ifo Business Climate Index for Germany is published. To be precise, following the need-to-know principle, only certain members of the EBDC team can transfer the raw data for further processing and publication from a protected internal ifo micro-database onto the dedicated internal EBDC server. It is hosted on ifo premises and in closed-shop mode under the 24/7-surveillance of ifo's IT department. Identifying or contact information of participating firms (e.g. company names, addresses etc.) is strictly separated from their survey answers. They are only temporarily combined for the generation of the panel data sets, consistency checks, and record linkage. EBDC guest users never get access to this raw data or any other identifying firm information. The term "raw" refers to the fact that this new data is not yet fully documented, variables and values are not labeled, variable (names) are not standardized over time, etc. All these tasks and many more are performed by the EBDC team. Additionally, the data is further pseudonymized. To this end, among other things, the remaining location information (i.e. the federal state variable) – is deleted for large firms to ensure their anonymity. Furthermore, the EBDC team conducts the machine learning-based record linking procedure of the ifo survey data with the external company data bases, but deletes any identifying information afterward as described above.

Despite anonymization and pseudonymization as well as a data retention period, to safeguard the identity of the firm survey participants and the confidentiality of their survey answers, guest researchers can only work at protected, supervised, and well-equipped guest researcher workstations in the isolated premises of the EBDC within the ifo Institute. Identities of all first-time guest researchers are checked by the EBDC team before granting access to the data. During visits by guest researchers, the EBDC team is always present and available to provide assistance and answer questions. Each guest researcher is assigned a password-protected user account and working folder to which only he/she has access. Data is stored on a dedicated and encrypted EBDC server run in closed-shop mode and backed up regularly. Any direct data import or extraction (through the internet or physical storage media) by guest researchers is technically disabled and contractually prohibited. Therefore, all data imports to the working folder (e.g. prepared statistical program code or additional external data) and exports of research results from the workstations have to be performed by the EBDC team. Only aggregated results can be exported for guest researchers (e.g. regression tables, figures, etc.). Furthermore, the EBDC data experts review all results to ensure that guest researchers comply with the contractual prohibition on re-identification and that individual firms cannot be identified (intentionally or accidentally). In addition, the EBDC team is happy to provide

[15] Personalized data of guest researchers (e.g. from their applications etc.) is of course compiled and stored securely, GDPR-compliant, and is only accessed for the intended purposes as well.
[16] As of early 2023, the EBDC does not conduct surveys of its own, but receives the data from the respective ifo Centers (e.g. the ifo Business Survey from the ifo Center for Macroeconomics and Surveys).

in-depth advice on the ifo micro-data sets, since they have years of experience in creating them.

These and other technical, organizational, and contractual security measures are specified in the ifo/EBDC IT security concept as well as the data protection concept. The EBDC team reviews these documents regularly with ifo's legal counsel, ifo's data protection officer as well as ifo's head of IT to ensure the highest level of confidentiality and security of the data sets.

6.4 FAIR Data

Most of ifo's research data sets published regularly by the EBDC are created in a way to fulfill the so-called FAIR principles (Findable, Accessible, Interoperable, Reusable). The EBDC creates standardized metadata and assigns an unique persistent digital object identifier (DOI) to them. Hence the data can be easily found via all common (data) search engines on the Internet.[17] Furthermore, the EBDC has a standardized process in place to access the data (see Section 6.5) and the documentation provided by the ifo Centers as well as the EBDC team (e.g. this handbook, variable lists, questionnaires) enables guest researchers to familiarize themselves with the data quickly. The data sets are usually provided in DTA format for Stata, but it can be converted easily. Lastly, the whole purpose of the EBDC is to make ifo's surveys reusable. Guest researchers themselves can make their analyses reusable if they decide to archive them at the EBDC as well (see Section 6.6).

6.5 Data Access

In general, the EBDC offers non-discriminatory access to its datasets (and several other services) free of charge. Most data sets are published regularly, but the periodicity of data publication and the length of the retention period depends on the particular data set, as described above. Guest researchers with an affiliation to a research institution or university may use the ifo micro-data sets during a short-term research stay at the EBDC in Munich for strictly academic, non-commercial research projects.[18] Guest researchers have to submit an application form including a short project description) and acknowledge EBDC's terms and conditions. The application is reviewed and approved (or in rare cases declined) by the head of the EBDC and by the EBDC Board in case of special requests. In addition, since the number of secure workstations is limited, all visits to the EBDC have to be scheduled in advance. In

[17] As a member of the Dara consortium (`https://www.da-ra.de`), the EBDC can assign its own DOIs (e.g. `https://doi.org/10.7805/ebdc-ibs-ind-2022a` for the ifo Business Survey in the manufacturing industry of 2022).

[18] Only few ifo data sets such as the ifo Education Survey are available as scientific use files and will be sent to researchers after (granted) application. More, up-to-date information about access regulations for each data set provided by the EBDC as well as the application form and EBDC's terms and conditions of use can be found on the EBDC webpage: `www.ifo.de/en/ebdc`.

addition, researchers are obliged to cite the micro-data appropriately and to provide a copy of their publication(s) free of charge, unsolicited, and promptly after publication. As of early 2023, data access is restricted to the secure workstations at the EBDC (and in some cases to remote execution by the EBDC team). However, the EBDC team is working on several secure interactive remote access solutions to simplify data access, especially for international guest researchers.

6.6 The EBDC as a Data Repository

As an accredited research data center, the EBDC acts and is listed as a trusted research data repository in the Registry of Research Data Repositories (Re3data), which is endorsed i.a. by the American Economic Association and for EU Horizon projects.[19] Research results from EBDC guest researchers based on ifo/EBDC data sets can be safely archived on a long-term basis and, if requested, published for replication and/or secondary analysis in the secure environment of the EBDC. In this way and as an institutional member of the LMU Open Science Center, the EBDC directly supports the implementation of the principles of good scientific practice, promotes the idea of Open Science and thus increases the trustworthiness, transparency, and efficiency of the research process.[20]

If replication data for research projects conducted at the EBDC are to be archived and/or published, all data must be submitted by the guest researcher to the EBDC team in a structured form with documentation and metadata. The EBDC team is happy to guide researchers in preparing the necessary documents and provides advice on data preparation according to the FAIR principles. However, the documentation itself has to be done by the researchers. The EBDC checks the completeness of a data package to be archived (data set, documentation, and program code). Undocumented, unstructured "data dumps" are not accepted for archiving. Furthermore, the EBDC team does not check the scientific validity of the methods and results. The EBDC can register the standardized metadata, add an appropriate usage license and assign a persistent digital object identifier (DOI) for these data packages, so that the data can be easily found via all common (data) search engines on the Internet. Depending on the sensitivity of the data, data access for secondary users can be defined as download, provision on request, or data processing only at the secure workstations in the EBDC.

A Data Archiving Agreement and Data Transfer Protocol regulate the details regarding archiving/publication between primary researchers and the EBDC. Among other things, the method of publication, the access concept, as well as the duration of archiving, are determined in the

[19] See www.re3data.org, https://social-science-data-editors.github.io/guidance/Requested_information_hosting.html#list-of-additional-acceptable-trusted-repositories-in-economics, and https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

[20] See https://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html and https://www.osc.uni-muenchen.de/about_us/index.html.

agreement. Secondary researchers who wish to use the data deposited at the EBDC have to apply for the data and accept EBDC's terms and conditions.

6.7 The Research Data Center of the Bavarian State Office for Statistics in the EBDC

In addition to survey data from research institutes, such as the ifo Institute, and the private sector, high-quality micro-data from official statistics are a vital source for empirical economic research and evidence-based policy advice. For this reason, the ifo Institute and the Bavarian State Office for Statistics have been cooperating since 2013 and have created a branch office of the Research Data Center of the Statistical Offices of the Länder at the EBDC, which is administratively and spatially separated from the rest of the EBDC. Upon application and by appointment, guest researchers from academic institutions can work there under supervision at two secure guest workstations.[21] Over 100 formally anonymized micro-data sets from all areas of life are available there. These data sets have greater information content than factual anonymized Scientific Use Files, which are typically sent to researchers directly.

[21] More information about the Bavarian Statistical Office and its Research Data Center can be found here: `https://www.statistik.bayern.de/service/forschungsdatenzentrum/index.html` (German only). Further Information (in English) about specific data sets, the application process, etc. can be found here: `https://www.forschungsdatenzentrum.de/en`.

Figure 6.1: Organization chart of the EBDC