Appendix, "Data, Intangible Capital, and Productivity," by Carol Corrado, Jonathan Haskel, Massimiliano Iommi, Cecilia Jona-Lasinio, Filippo Bontadini. In Technology, Productivity, and Economic Growth, Studies in Income and Wealth volume 83, edited by Susanto Basu, Lucy Eldridge, John Haltiwanger, and Erich Strassner. Chicago, IL: University of Chicago Press, 2025.

This appendix provides a description of the cost-based approach and data sources used to estimate market sector investment and capital stocks in data stores and data intelligence (the components of data not currently included in official national accounts), databases and computer software introduced in this paper. The methodological approach is the same adopted to estimate own-account brand, design, organizational capital, and new financial products in EUKLEMS & INTANProd database (Bontadini et al. 2023). The data sources and main steps of the calculation are described below. A final section provides an illustration of the approach followed to estimating domestically sourced component of intangible investment.

Data assets measurement

The method adopted to generate estimates of data assets follows a cost-based approach assuming that the value of an asset can be obtained as the sum of the costs sustained for producing it. The basic approach can be summarized as follows:

$$Y^i_{bc} = COMP^i + IC^i + CK^i + T^i \qquad (A1)$$

where i = asset type, Y is the value of the produced asset at basic prices, $COMP^i$ is the labor cost of the relevant personnel measured as compensation of employees, IC^i are intermediate costs related to the activity, CK^i refers to the costs of capital services and T^i to net taxes on production related to these activities. But notice that besides $COMP^i$ the remaining components in equation (A1) are not directly measurable, thus the sum of these unmeasurable components, set equal to α , is a factor that must be approximated. So equation (A1) can be re-written as:

$$Y^{i}_{bc} = COMP^{i} + \alpha^{i} \qquad (A2)$$

where, notice that Y^{i}_{bc} can be measured directly by computing the compensation of employees (COMP) and finding a proxy for α. Thus COMP can be obtained as:

$$COMP^{i} = EMP^{i}_{tot} * W^{i}_{avg} * t^{i} \qquad (A3)$$

where EMP^{i}_{tot} is the total number of employees employed for producing the relevant asset, W^{i}_{avg} is the average remuneration (average wage) and t^{i} refers to the time spent on these activities (table A1 below shows the time assumptions (t) underlying the calculations developed in this paper). Using equation (A3) and substituting it in equation (A2) where it is assumed that $\alpha^{i}=COMP^{i}*bp^{i} - COMP^{i}$, the value of the produced asset is determined as:

$$Y^{i}_{bc} = COMP^{i}*bp^{i} \qquad (A4)$$

where *bp* is a blow-up factor that accounts for other cost components besides the compensation of employees and essential to develop a measure of output consistent with national accounts. The blow-up factors for each asset are measured using the ratio of gross output (GO) over the compensation (COMP) of all persons engaged where GO is adjusted to exclude national accounts own-account intangibles and intermediate purchases of intangibles that are capitalized in our framework. COMP is compensation of employees plus an estimate of compensation of self-employed. This adjustment is relevant, especially for those industries producing a sizable amount of intangibles whose production structure is assumed rather similar to the internal intangible factory described in the main text. The blow-up factors are estimated at the detailed industry level using US supply and use tables. The bp of the relevant industries averaged over 1997-2020 are then applied to each data asset.[1] In this paper, the blow-up factors are set equal to 1.7 for data intelligence and 1.8 for the other assets.

Main sources

The estimates of data assets illustrated in this paper have been produced applying equation (A4) across industries and countries. The main information needed to implement the calculation for each individual data asset is the following: i) a detailed list of occupations engaged in producing data assets; ii) occupation-specific (and industry-specific, if relevant) assumptions on the share of time spent in producing each data asset; iii) data on the number of employees for the relevant occupations and their compensations; iv) blow-up factors to account for other cost components (intermediate consumption and gross operating surplus) to derive an output measure consistent with national accounts definitions.

More specifically, table A1 shows the list of occupations that are assumed to be engaged in producing data assets and computer software capital formation based on ISCO-08 codes and of the time-use assumptions by asset as detailed below.

Notice that the selection of relevant occupations is constrained by the level of

---

[1] For data stores, we apply the average of blow-up factors calculated for Miscellaneous professional, scientific, and technical services, Computer systems design and related services, and Publishing industries, except internet (includes software); for databases and software, the average of Computer systems design and related services, and Publishing industries, except internet (includes software); for data intelligence, the average of Management of companies and Miscellaneous professional, scientific, and technical services.

detail of available data sources. For this paper, we use micro-data from the EU Structure of Earnings Survey (SES) for the years 2010, 2014, and 2018 and the EU Labour Force Survey (LFS) for 2008 and 2019. The SES provides information on the number of employees by occupation (at the three-digit level of the 2008 International Standard Classification of Occupations, ISCO) and economic activity and their annual earnings. The LFS, instead, provides data on employment with no information on wages. In the LFS, occupations are available at three-digit level of ISCO classification for all countries. On the other hand, the SES provides data for 11 countries at two-digit level, that have to be further expanded to three-digit level to get a coherent information set. The higher level of disaggregation for the SES wages from two to three digits has been obtained computing the share of relevant three-digit occupation from the LFS then applied to the SES variables.

Estimating the time-use factors

In this paper, we make a further step in developing the estimates of the time use factors compared to previous literature. In particular, we generate our measures resorting to a very high level of disaggregation of the occupation classification, four-digit level, for a group of European economies.[2] First we have identified the relevant occupations and the

[2] Ideally, the selection of the occupations and the definition of the corresponding time-use assumptions would have to be done at four digits level in the ISCO classification. But the available sources, the LFS and SES provide data at three digits level thus requiring a careful analysis of each code to identify the best time use assumption. As a matter of fact, not all the occupations within each ISCO code devote their working time to data production to the same extent or to produce the same type of asset.
So far, approaches to measure data assets dealt with aggregate occupation groups tweaking the assumption on the time-use factors to consider that each minor group includes workers who differ in how much time they spend producing a given data asset (or who do not create data assets at all). However, this approach adds a further layer of imprecision to estimates which are in themself based on many assumptions.

corresponding time-use assumptions at four-digit level of the US Standard occupations Classification (SOC 2010) and then we have used the data from the US Occupational Employment and Wage Statistics 2019 (OEWS) to compute a set of weights to generate an estimate of the time-use factor aggregated at 3-digit ISCO level.

More precisely, the approach for measuring the time-use factor for each data asset can be summarized as follows: i) Identification of the ISCO unit groups included in each ISCO minor group;[3] ii) for each ISCO unit group, identification of the corresponding occupation in the 2010 SOC (based on a crosswalk between the ISCO-08 and the 2010 SOC available from the US Bureau of Labor Statistics); iii) assignment of a time-use factor to each four-digits SOC occupation.

Finally, the time use factors at three digit level have been computed as the weighted average of the time-use factor of the corresponding four-digits SOC where the weights have been generated from employment data gathered from the 2019 OEWS.[4]

Capital stock and price deflators

The estimates of capital stock in real terms used in the econometric analysis are generated applying the perpetual inventory method (PIM) based on the aggregation of real investment over time allowing for declines in efficiency and value until the assets reach the end of their service lives and are retired.[5]

---

[3] In the ISCO classification, minor groups are occupations defined at the level of three digits and unit groups at the level of four digits.

[4] However, this approach has some drawbacks. It is implicitly assumed that the employment shares of four-digit occupations within each three-digit group are the same in all European countries and equal to those in the US. In addition, the aggregate percentages are used to develop industry-level calculations.

[5] See Hulten Charles R. "The Measurement of Capital" in Ernst R. Berndt and Jack E.

We assume economic depreciation is geometric, in which case the real stock of data asset j in industry i at the end of year t ($Kq_{i,tj}$) is defined as:

$$Kq_{i,tj} = Kq_{i,t-1j} * (1-\delta_j) + Iq_{i,tj} \qquad \text{(A5)}$$

where $Kq_{i,t-1j}$ is the real stock of asset j in industry i at the end of year t-1, δ_j is the annual depreciation rate for asset j and $Iq_{i,tj}$ is real investment for asset j in industry i during year t. Note that depreciation rates are asset-specific and are assumed to not vary across industries and over time.

Real investment in each type of data asset is obtained by dividing its nominal investment flow by an appropriate price index. The real value for data intelligence has been computed applying the deflator of non-national accounts intangibles while for data stores, computer software and databases exploiting the harmonized software deflator developed for the analytical module of the EUKLEMS & INTANProd databases (see Bontadini et al. 2023).

For what concern the depreciation rates, for data intelligence it is the average of the depreciation rates of non-national accounts intangible assets (0.35), while for the other data assets it is the same depreciation rate used for software in EUKLEMS & INTANProd (0.315).

Estimates of data assets have been developed by industry, at the level of Nace sections, and then aggregated at the market sector level, defined as all industries excluding Nace sections L (real estate activities), O (public administration and defense;

Triplett, eds., Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth, Chicago, University of Chicago Press, 1991.

compulsory social security), P (education), Q (human health and social work activities), T (activities of households as employers; undifferentiated goods - and services-producing activities of households for own use).

Domestic Component of Intangible Investment: Data Sources and Estimation Method

In what follows we illustrate the main steps for estimating domestically sourced component of intangible investment.

Measures of domestically produced investment in R&D and computer software and databases are obtained using data from national supply and use tables. First, it is necessary to compute the share of gross output in total resources for domestic use, SGOD, as follows:

$$SGOD_i = (GO_i - EX_i) / (GO_i - EX_i + IM_i) \qquad \text{(A6)}$$

where GO_i, EX_i and IM_i are gross output, exports and import of product i (i= CPA_M72 for R&D, and CPA_J62-63 for computer software and databases).

Then, the domestic component of each asset is generated multiplying national accounts investment by the corresponding share of gross output in total resources for domestic use. A main assumption underlying this approach is that the share of the domestic component is the same across different uses (intermediate consumption, final consumption, and investment).

For those intangible assets not included in national accounts (brand, design and organizational capital), the domestic component is computed as the sum of the own-account investment and an estimate of the domestically sourced purchased component. That is, for organizational capital:

$$I_{dom}^{OrgCap} = (I^{OrgCap(OA)} + I_{dom}^{OrgCap(Purchased)}) \quad (A7)$$

New financial products are only domestically produced, and that there are no available data sources for estimating imported training. The imported component for training is deemed very small and is ignored.

To estimate equation (A7) it is essential to measure the domestically sourced purchased component of brand, design, and organizational capital. Estimates of these components are generated from the information gathered from the world input-output tables reporting the intermediate use of domestic output and intermediate use of imports from other countries disaggregated by product, for each industry in each country (World Input-Output Database, or WIOD, available at https://www.rug.nl/ggdc/valuechain/wiod/?lang=en).

Resorting to these data, it is possible to compute the share of domestic output in market sector intermediate consumption of the following products: advertising and market research services (CPA M73), architectural and engineering services, technical testing, and analysis services (CPA M71) and legal and accounting services, services of head offices and management consulting services (CPA_M69_70). Then the domestically sourced purchased component of brand, design and organizational capital can be generated by multiplying the purchased investment component by the share of domestic output in total intermediate consumption for the relevant products listed before (CPA M73 for brand, CPA_M71 for design, and CPA_M69_70 for organizational capital).

The 2016 WIOD release provides an annual time-series of world input-output tables from 2000 to 2014. For the most recent years, shares have been extrapolated regressing the 2000-2014 shares on a linear time trend.

Correlation tables

Tables A2 A) and B) below shows the correlations between the estimated 2010 to 2019 time series for the value-added shares of data assets and intangibles and for the rate of growth of the shares.

Table A1. Relevant Occupations for defining Time-use Assumptions to estimate Data Assets and Computer Software

ISCO-08 sub-major group	ISCO-08 minor group	Occupation description	Time-use (%)			
			Software	Databases	Data Stores	Data Intelligence
1	**1**	**Managers**				
11	**11**	**Chief executives, senior officials and legislators**				
111	11	Legislators and senior officials	0.00	0.00	0.00	0.00
112	11	Managing directors and chief executives	0.00	0.00	0.00	0.17
12	**12**	**Administrative and commercial managers**				
121	12	Business services and administration managers	0.00	0.00	0.00	0.00
122	12	Sales, marketing and development managers	0.00	0.00	0.00	0.10
13	**13**	**Production and specialized services managers**				
131	13	Production managers in agriculture, forestry and fisheries	0.00	0.00	0.00	0.00
132	13	Manufacturing, mining, construction, and distribution managers	0.00	0.00	0.00	0.00
133	13	Information and communications technology service managers	0.00	0.10	0.10	0.10
134	13	Professional services managers	0.00	0.00	0.00	0.00
14	**14**	**Hospitality, retail and other services managers**				
141	14	Hotel and restaurant managers	0.00	0.00	0.00	0.00
142	14	Retail and wholesale trade managers	0.00	0.00	0.00	0.00
143	14	Other Services Managers	0.00	0.00	0.00	0.00
21	**21**	**Science and engineering professionals**				
211	21	Physical and earth science professionals	0.00	0.00	0.10	0.40
212	21	Mathematicians, actuaries and statisticians	0.05	0.06	0.13	0.32

213	21	Life science professionals	0.00	0.02	0.11	0.28
214	21	Engineering professionals (excluding electrotechnology)	0.00	0.00	0.10	0.25
215	21	Electrotechnology engineers	0.00	0.00	0.10	0.36
216	21	Architects, planners, surveyors and designers	0.00	0.00	0.18	0.09
24	**24**	**Business and administration professionals**				
241	24	Finance professionals	0.00	0.00	0.10	0.26
242	24	Administration professionals	0.00	0.07	0.07	0.11
243	24	Sales, marketing and public relations professionals	0.00	0.00	0.15	0.21
25	**25**	**Information and communications technology professionals**				
251	25	Software and applications developers and analysts	0.48	0.10	0.03	0.01
252	25	Database and network professionals	0.10	0.47	0.10	0.03
263	26	Social and Religious Professionals	0.00	0.00	0.00	0.03
33	**33**	**Business and administration associate professionals**				
331	33	Financial and mathematical associate professionals	0.00	0.00	0.10	0.25
332	33	Sales and purchasing agents and brokers	0.00	0.00	0.00	0.00
333	33	Business services agents	0.00	0.02	0.02	0.04
334	33	Administrative and specialized secretaries	0.00	0.00	0.00	0.00
335	33	Regulatory government associate professionals	0.00	0.00	0.00	0.02
35	**35**	**Information and communications technicians**				
351	35	Information and communications technology operations and user support tec	0.01	0.02	0.03	0.04
352	35	Telecommunications and broadcasting technicians	0.00	0.00	0.10	0.05
41	**41**	**General and keyboard clerks**	0.00	0.00	0.07	0.00

42	**42**	**Customer services clerks**				
422	42	Client information workers	0.00	0.00	0.05	0.00
43	**43**	**Numerical and material recording clerks**				
431	43	Numerical clerks	0.00	0.00	0.16	0.00
432	43	Material-recording and transport clerks	0.00	0.00	0.05	0.00

Note: 413 includes data entry clerks (4132); 422 includes survey and market research interviewers (4227); 431 includes statistical, finance and insurance clerks (4312).

Table A2. Correlation of estimates

A. Investment shares of adjusted value added

	Data capital	Data stores	Databases	Data intelligence	**Software**	**Intangible capital**	Digitized information	Innovative property	Economic competencies	**Data-intensive intangibles**	Organization	Design	Brand	R&D
Data capital	1													
Data stores	0.746***	1												
Databases	0.663***	0.594***	1											
Data intelligence	0.905***	0.442***	0.369***	1										
Software	0.795***	0.284**	0.526***	0.859***	1									

Intangible capital	0.553***	0.0476	0.461***	0.634***	0.866***	1								
Digitized information	0.111	-0.263*	0.377***	0.156	0.577***	0.763***	1							
Innovative property	0.247*	0.0102	0.222*	0.283**	0.389***	0.632***	0.368***	1						
Economic competencies	0.678***	0.244*	0.422***	0.740***	0.836***	0.791***	0.478***	0.101	1					
Data-intensive intangibles	0.643***	0.140	0.410***	0.741***	0.900***	0.921***	0.602***	0.367***	0.943***	1				
Organization	0.576***	0.140	0.456***	0.627***	0.838***	0.864***	0.673***	0.210*	0.946***	0.937***	1			
Design	0.142	-0.121	0.0827	0.235*	0.369***	0.577***	0.369***	0.904***	0.0825	0.386***	0.198	1		
Brand	0.610***	0.366***	0.242*	0.646***	0.554***	0.399***	0.0233	-0.176	0.782***	0.667***	0.557***	-0.188	1	
Memo: R&D	0.244*	0.0824	0.317**	0.214*	0.328**	0.601***	0.364***	0.943***	0.0998	0.309**	0.219*	0.726***	-0.193	1

B. Growth rate of Investment shares of adjusted value added

	Data capital	Data stores	Databases	Data intelligence	**Software**	**Intangible capital**	Digitized information	Innovative property	Economic competencies	**Data-intensive intangibles**	Organization	Design	Brand	R&D
Data capital	1													
Data stores	0.834***	1												
Databases	0.746***	0.419***	1											
Data intelligence	0.919***	0.577***	0.729***	1										
Software	0.770***	0.528***	0.710***	0.778***	1									

Intangible capital	0.0963	0.0784	0.0609	0.0910	0.102	1								
Digitized information	-0.0361	-0.0755	-0.00605	-0.0114	0.0291	0.750***	1							
Innovative property	-0.0917	-0.0748	-0.0873	-0.0803	-0.0350	0.918***	0.622***	1						
Economic competencies	0.552***	0.518***	0.373***	0.472***	0.368***	0.364***	0.135	0.122	1					
Data-intensive intangibles	0.622***	0.581***	0.378***	0.546***	0.418***	0.320**	0.0391	0.116	0.904***	1				
Organization	0.409***	0.331**	0.340**	0.351***	0.370***	0.330**	0.180	0.106	0.785***	0.766***	1			
Design	0.350***	0.403***	0.148	0.262*	0.161	0.0432	-0.204	0.0136	0.247*	0.598***	0.301**	1		
Brand	0.494***	0.492***	0.233*	0.440***	0.268*	0.267*	0.0769	0.0996	0.844***	0.732***	0.379***	0.0905	1	
Memo: R&D	-0.109	-0.0913	-0.105	-0.0918	-0.0349	0.894***	0.631***	0.988***	0.114	0.0769	0.0743	-0.0721	0.120	1

Notes: ** $p<0.01$, *** $p<0.001$. Countries include Denmark (DK), Germany (DE), Finland (FI), France (FR), Italy (IT), Netherlands (NL), Spain (ES), Sweden (SE) and the United Kingdom (UK). Gross value added is adjusted to include all intangibles, as reported in the EUKLEMS & INTANProd database (analytical module).