

Online Appendix

A Data Appendix

A.1 Selection of utility patents

Intellectual property offices publish many documents that are called “patents”. For example, the USPTO publishes “plant patents”, “defensive publications”, “reissue patents”. To select the patents that most closely match the idea of *patents of invention* and to avoid double counting, we select documents based on the kind codes. Our goal is to mimic as closely as possible the concept of “first publication of granted patent”. We start with a list of kind code that corresponds to utility patents which we enumerate in Table A1. From this list, we make additional selections to remove non granted patents and to keep only first publications. Formally, we made the following selections:

- **United States:** We keep documents with kind code A (Granted patents prior to 2001), B1 (Granted patent published without an application), B2 (Granted patent published with an application, after 2001).
- **Germany:** We keep publications with kind code C (“*Patentschrift*”) or B (“*Auslegeschrift*”) without conditions. These corresponds to first publications of patents before 1970. After 1970, the publication process changed and a given patents could have several publication. We keep kind code A1 (“*Offenlegungsschrift*”) whenever a given patent (identified by the publication number without the kind code) has more than one publication. We then keep kind code C2 if the patent does not have a A1 publication but has more than one additional publication (on top of the C2). Finally, we keep all patents that have only one publication, except if this publication is a A1 document.
- **France:** We keep publications with kind code A and A5 without conditions (“*Brevet d’invention*”). We then keep kind code A1 (“*Demande de Brevet d’Invention*”) if there is only one publication for a given patent and if the publication year is earlier than 1971. Otherwise we keep publications with kind code A1 if there are more than one publication for the patent.
- **United Kingdom:** We keep documents with kind code A (Patent Application) if the publication number is lower than 2000000 or if the publication year is earlier than 1921. Otherwise, we keep A if there are more than one publication for the patent.

These rules are governed by the fact that the patent systems change over time. Typically in earlier years, all patent publications correspond to the one and only document that served as the final granted patent. In the most recent decades, patent offices published the patent applications along with other subsequent documents if the granting process was successful. Simply counting all patent applications would result in the inclusion of patents that have not been granted and to overestimate the number of patents in the most recent period. Note that we also release a version of the database where we did not make these restrictions and include all utilities patents, whose kind code are summarized in Table A1.

Table A1: GRANTED UTILITY PATENTS

Patent office	Time span (publication year)	Kind code(s)
DD	1950-1992	A, A1, A3, B
DE	1877-2013	A1, B, B3, C, C1, D1
FR	1902-2013	A, A1, A5*, B1*
GB	1893-2013	A, B*
US	1836-2013	A, B1*, B2*

Notes: The selected kind codes try to emulate the USPTO concept of "Granted Utility Patent". We restrict to the first publication or second publication without first publication kind codes in order to avoid double counting issues. We exclude patent *applications* and *revised* publications for the same reason. In the case of DD, we are limited by the availability of raw patent images and therefore include all types of publications. * indicates that the kind-code is considered only after 1980. This can be due to changes in the meaning of the kind-code or to its creation date.

A.2 Formats

The structure of a patent document can change over time as the patent office modernizes its publications and processes. We tracked these changes and adapted the statistical model that we used to each cases. Table A2 shows the different formats for each patent offices and the first and last patents of each format.

Table A2: PUBLICATION NUMBER AND PATENT FORMAT

Patent office	Publication number (range)	Format number	Earliest Year
DD	DD1 - DD123499	1	1951
DD	DD123500 -	2	1977
DE	DE1C - DE977922C	1	1877
DE	DE1000001B -	2	1957
FR	FR317502A - FR1569050A	1	1902
FR	FR1569051A - FR1605567A and then from FR2000001A1	2	1969
GB	GB189317126A - GB2000001A	1	1894
GB	GB2000001A -	2	1979
US	US1A - US1583766A	1	1836
US	US1583767A - US1920166A	2	1924
US	US1920167A - US3554066A	3	1933
US	US3554067A -	4	1971

Notes: The structure of a patent document can change over time. We track these changes and adapt the statistical model to each case. The table shows the different formats for each patent office and the associated first and last patents of each format. The earliest year corresponds to the first year in which the corresponding format was implemented. For more details, see [Annotation guidelines for the US, Germany, the UK and France](#).

A.3 Entities by country

In this Section, we detail the different types of entities matched for each country and what they usually means.

United States In the case of the US, the inventors and assignees are clearly separated entities. The inventor is the name of the person who conceived the invention while the

assignee is the entity (either a person, a firm, the government, a university...) who own the right of the patent. US patents also give information on the citizenship of patentees. In the case of inventors, this is the country of citizenship (e.g., “a citizen of the kingdom of Italy”) and in the case of assignee the legal origin of the firm when applicable (“a company duly organized under the laws of New Jersey”). Finally, the entity location gives the address of the inventor and assignee, usually at the city level. For more details, see the [Annotation guidelines for the US](#)

Germany In the case of Germany, inventors are referred to as “*Erfinder*” and assignees as “*Anmelder*”. Both entities can represent physical people while assignees can also be companies. Most of the patents filed before the 1950s do not include any inventor. Although it is likely that in that case, the inventor and the assignee can be the same person, we only label the entity as inventor when the term “*Erfinder*” is explicitly mentioned. German patents also give some information on the occupation of inventors or assignees from the denomination of their academic title (e.g., “*Dr.*”, “*Ing.*” or “*Pr.*”). Finally, the location is usually given by the city of the inventor or assignee. For more details, see the [Annotation guidelines for Germany](#) and [the specific guidelines for East-Germany](#)

France The case of France is similar to the case of Germany regarding inventors and assignees. Most of the patents have a “*déposant*” which we label as assignee while some patents also have an “*inventeur*” which we label as inventor. French patents do not give information on occupation or citizenship, except if extremely rare instances. The location is given at the county (“*département*” level in the case of a patentee located in France and at the country level for foreign inventors. For more details, see the [Annotation guidelines for France](#)

United Kingdom In the British case, the inventor and the assignees are not explicitly distinguishable. By convention, we denote each firm by an assignee and each person as an inventor. The British patents also include information on the occupation of the inventor, and in some case on the occupation of the assignee (e.g., “a clock manufacturing company”). Information on the citizenship of inventor and assignee are also provided like in the US. Finally, the location of the assignee and of the inventor is given as a full postal address. For more details, see the [Annotation guidelines for British patents](#).

A.4 Data coverage

This Section presents the coverage of each entities as a share of patents for the five patent offices considered. Precisely:

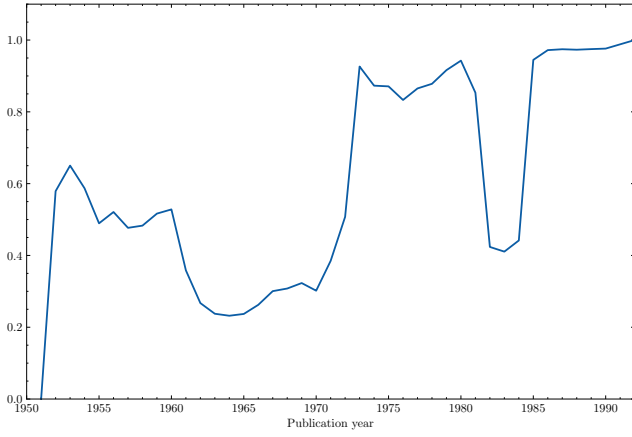
- Figure [A1](#) plots the yearly share of patents with at least one inventor
- Figure [A2](#) the yearly share of patents with at least one assignee
- Figure [A3](#) plots the yearly share of patents with at least one location
- Figure [A4](#) the share of patentees that are matched with a location

- Figure A5 plots the yearly share of inventors with at least one entity occupations
- Figures A6 plots the yearly share of inventors with at least one entity citizenship
- Figure A7 shows the relative share of each level of geographical matching.
- Figure A8 reports the composition of the geocoding by source: either using commercial geocoding supplier: HERE or GMAPS or manually

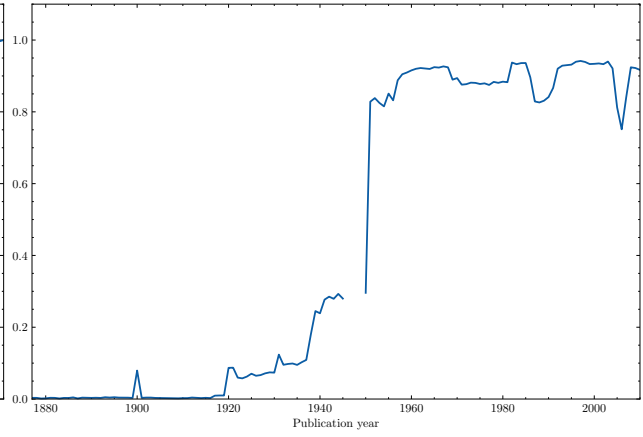
Finally, Figures A9, A10 report the number and share of patent publications by source: either from PatentCity, from de Rassenfosse et al., 2019 or from the expansion (that is, we expand the entities and relationships to all patents of the same family when information is missing). Figures A11 and A12 compare the coverage of the PatentCity database with the coverage of the Claims database that we take as the universe of patents.

Figure A1: SHARE OF PATENTS WITH AT LEAST ONE INVENTOR

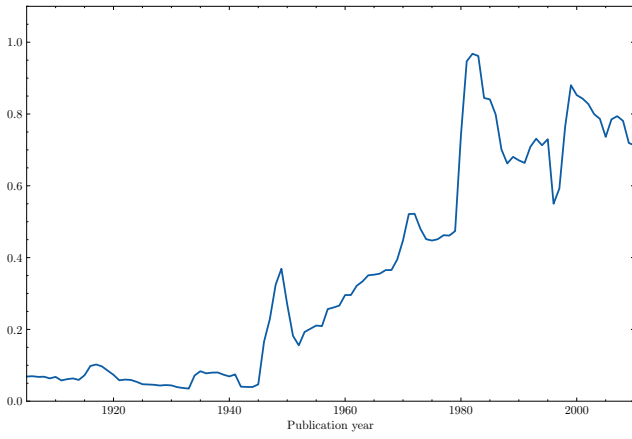
(a) DD



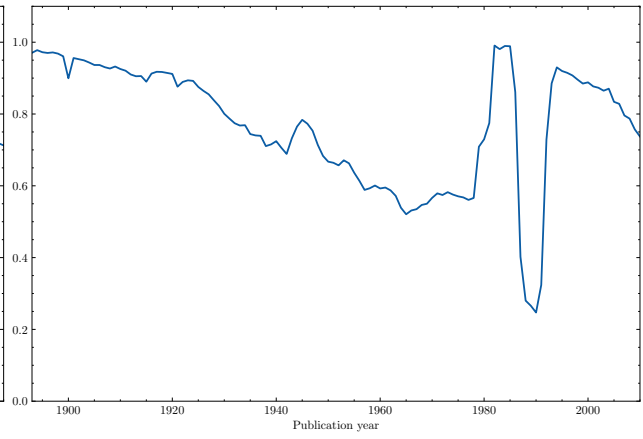
(b) DE



(c) FR



(d) GB



(e) US

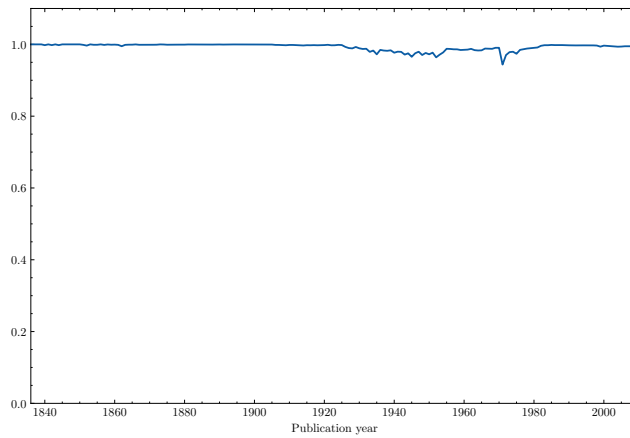
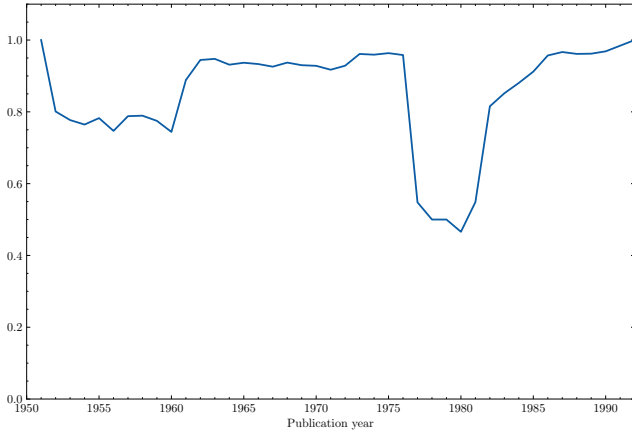
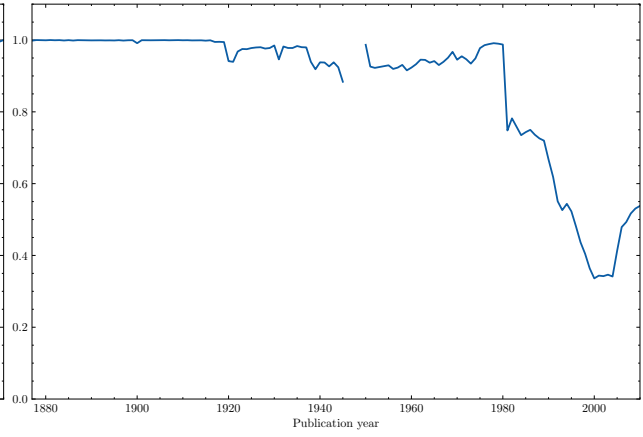


Figure A2: SHARE OF PATENTS WITH AT LEAST ONE ASSIGNEE

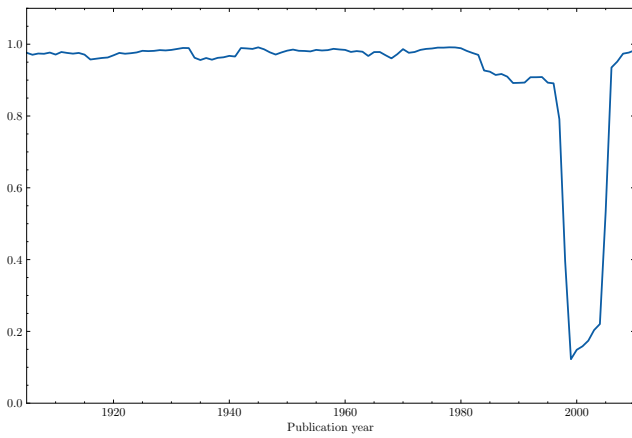
(a) DD



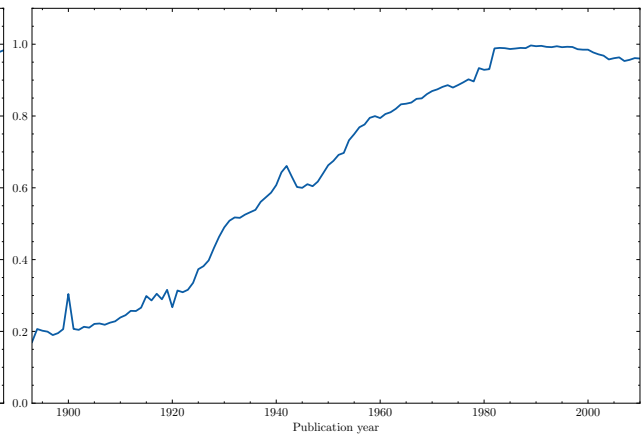
(b) DE



(c) FR



(d) GB



(e) US

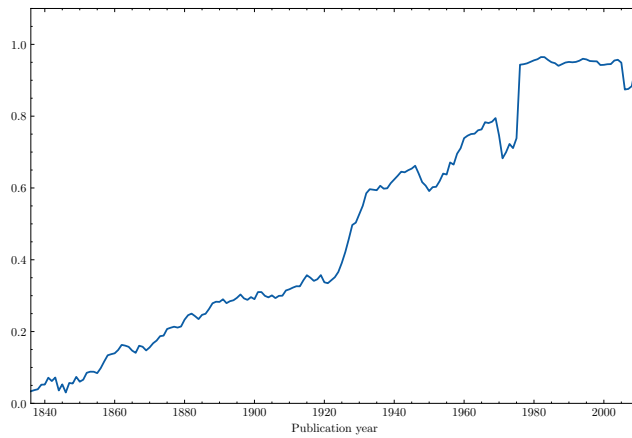
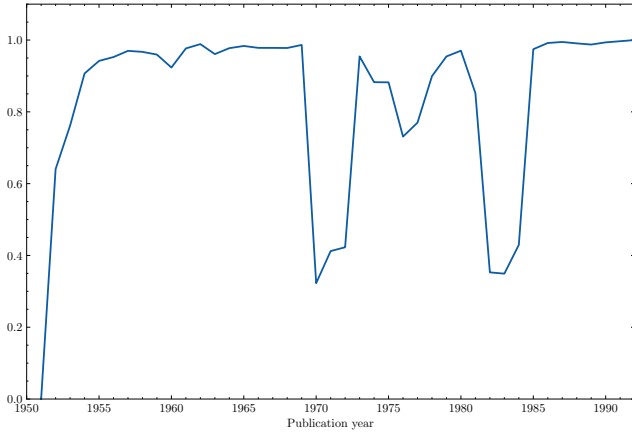
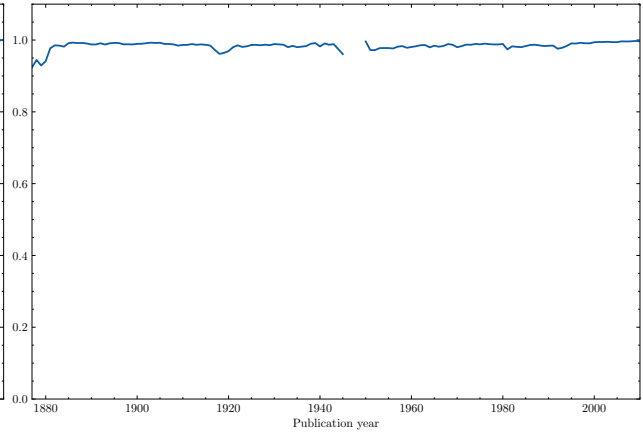


Figure A3: SHARE OF PATENTS WITH AT LEAST ONE LOCATION

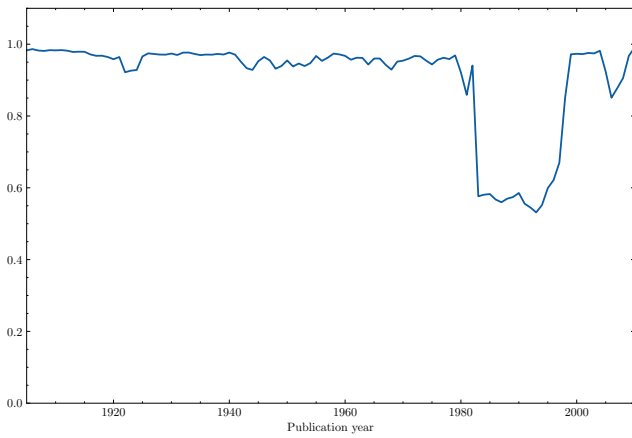
(a) DD



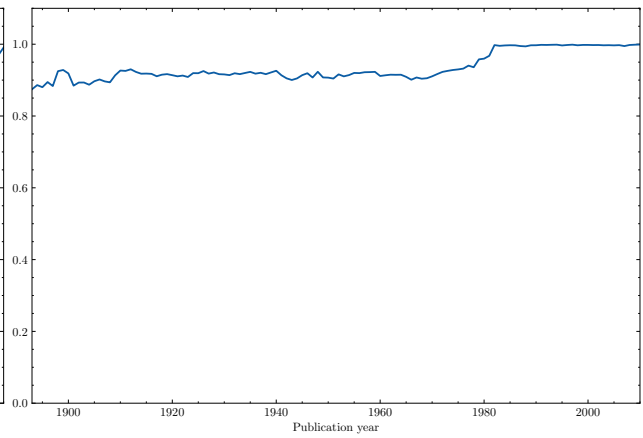
(b) DE



(c) FR



(d) GB



(e) US

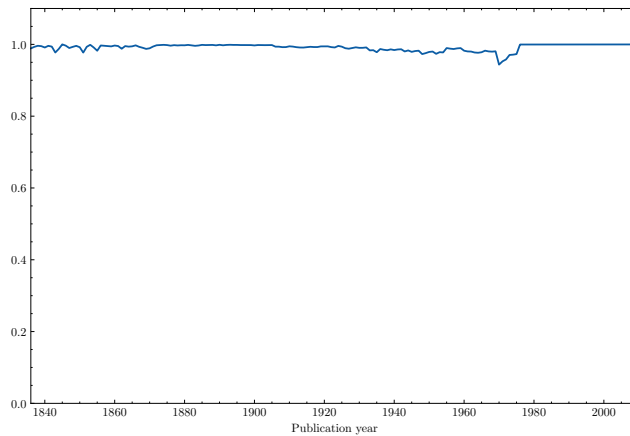
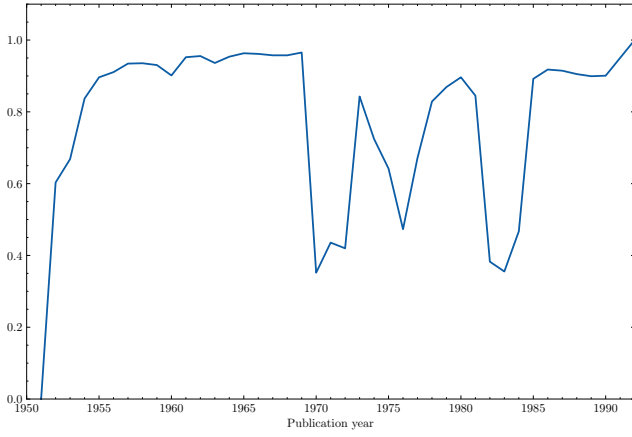
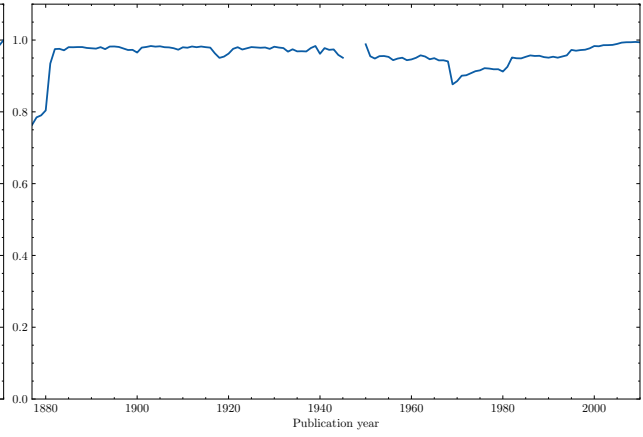


Figure A4: SHARE OF PATENTEES WITH A DETECTED LOCATION

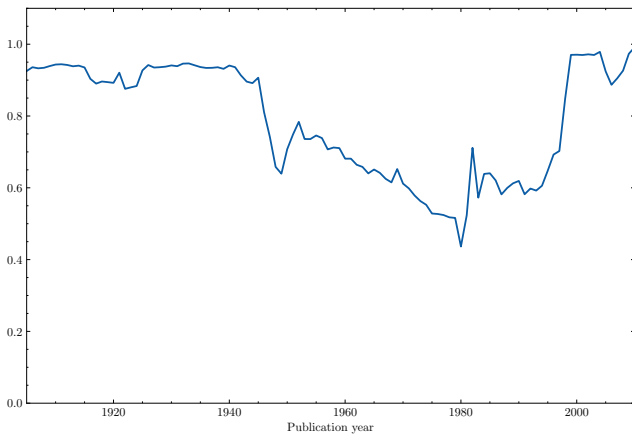
(a) DD



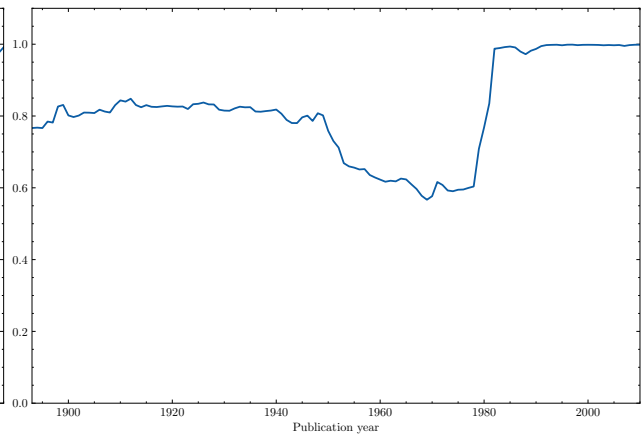
(b) DE



(c) FR



(d) GB



(e) US

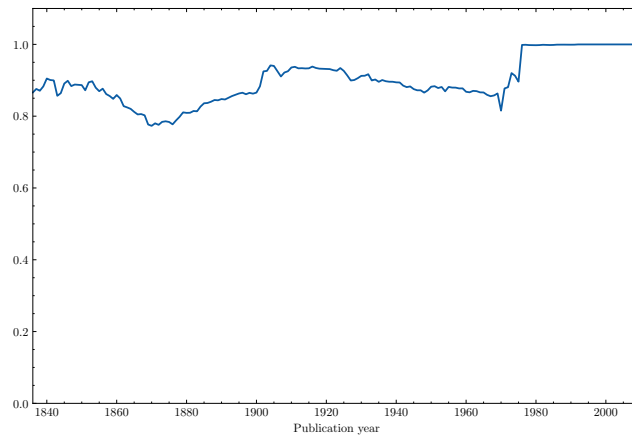


Figure A5: SHARE OF INVENTORS WITH A DETECTED OCCUPATION

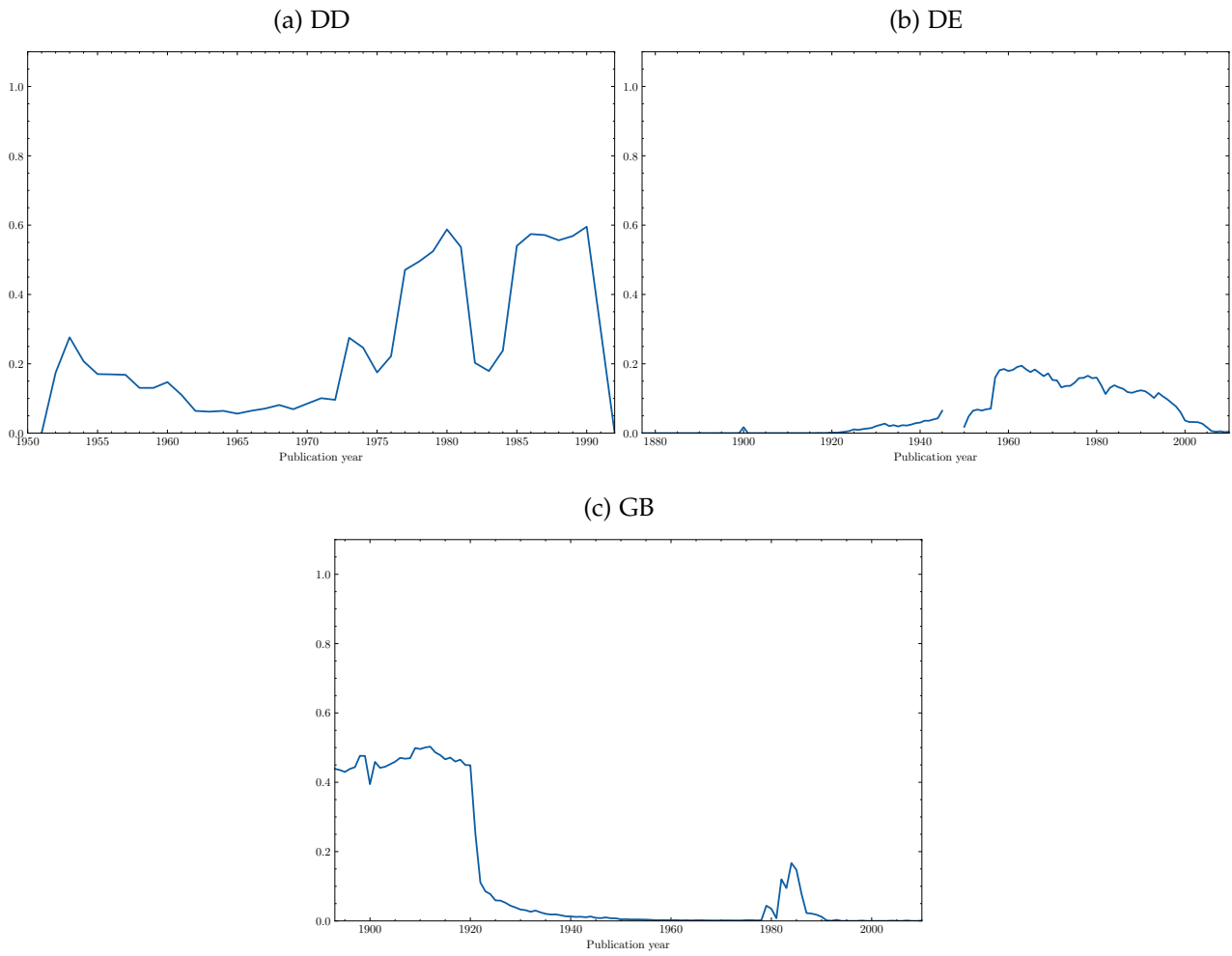


Figure A6: SHARE OF INVENTORS WITH A DETECTED CITIZENSHIP

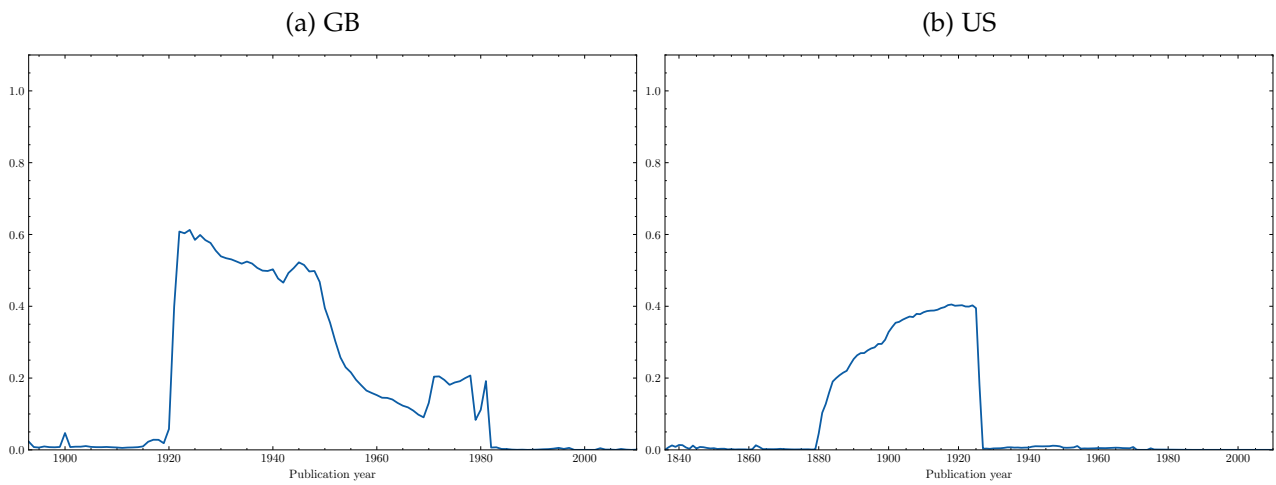


Figure A7: COMPOSITION OF THE MOST DETAILED LEVEL OF GEOCODING

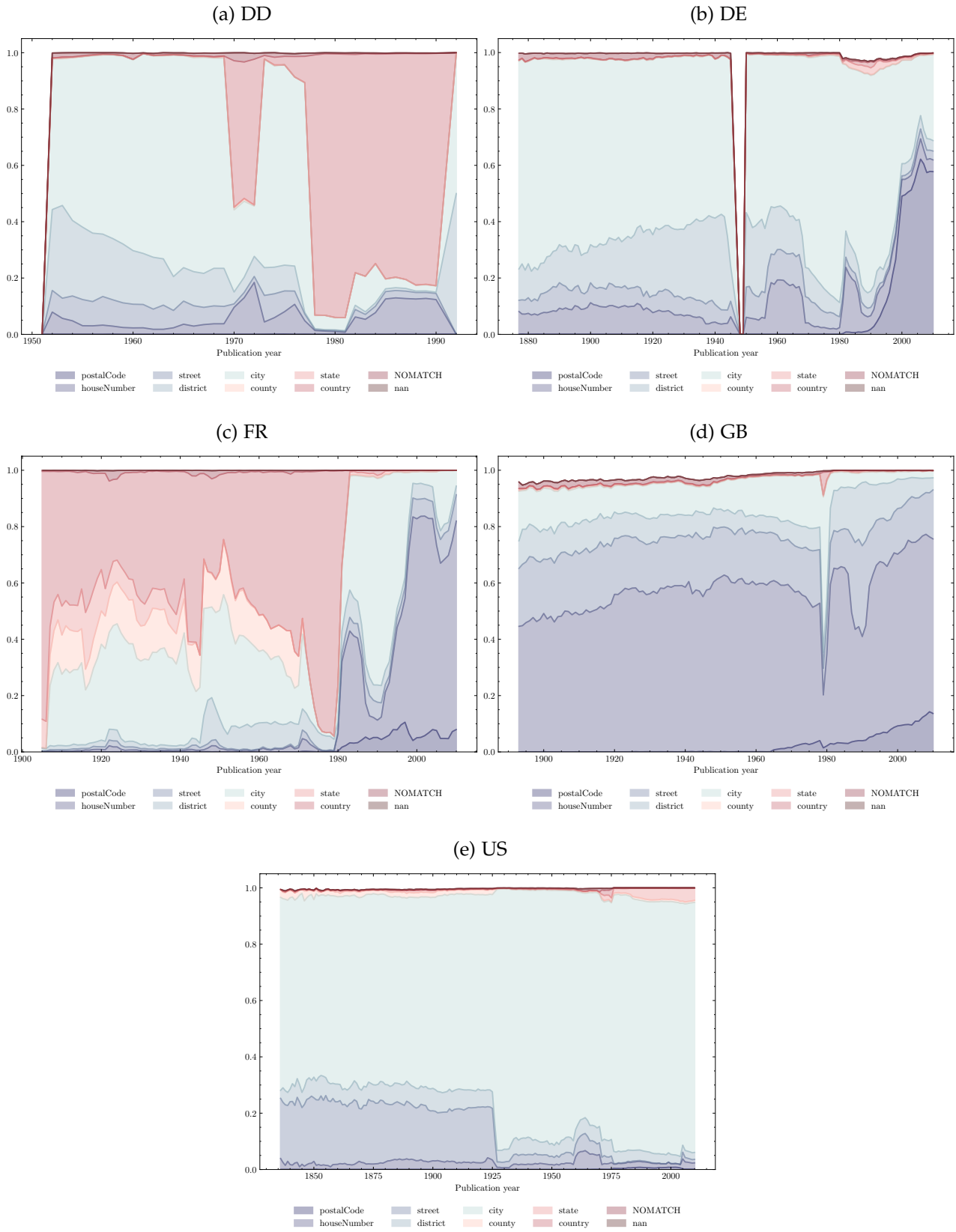


Figure A8: COMPOSITION OF GEOCODING BY GEOCODING SOURCE

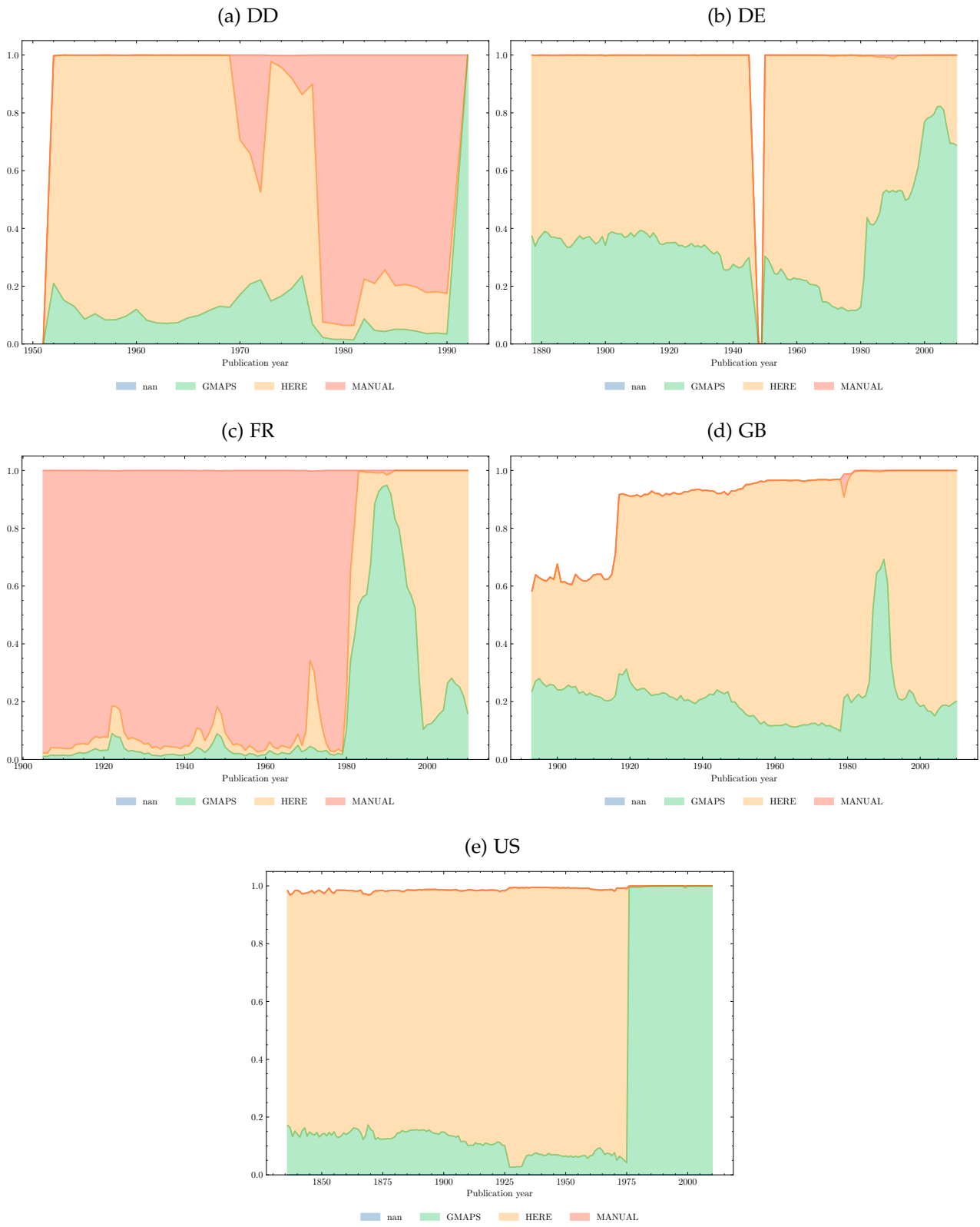
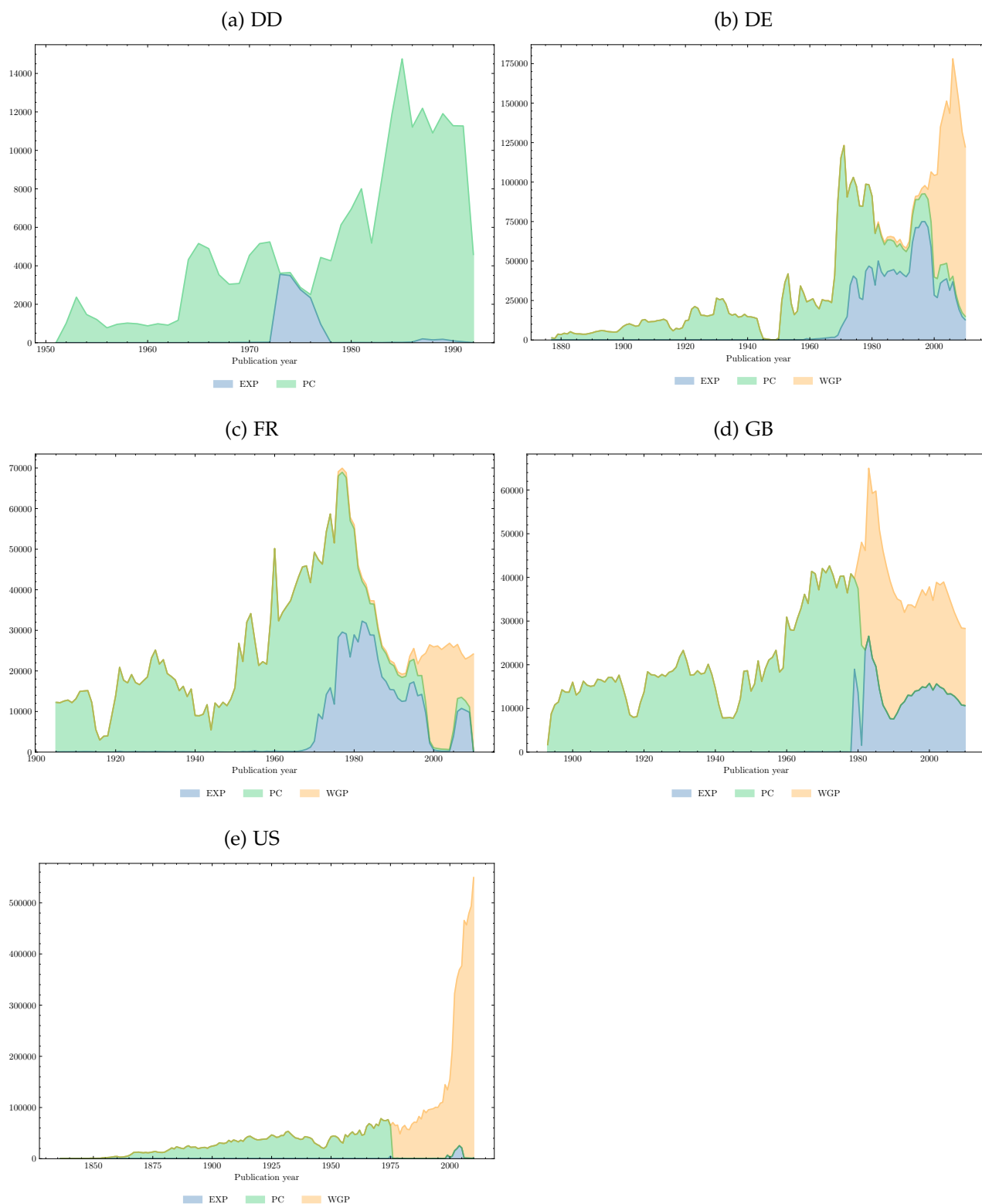
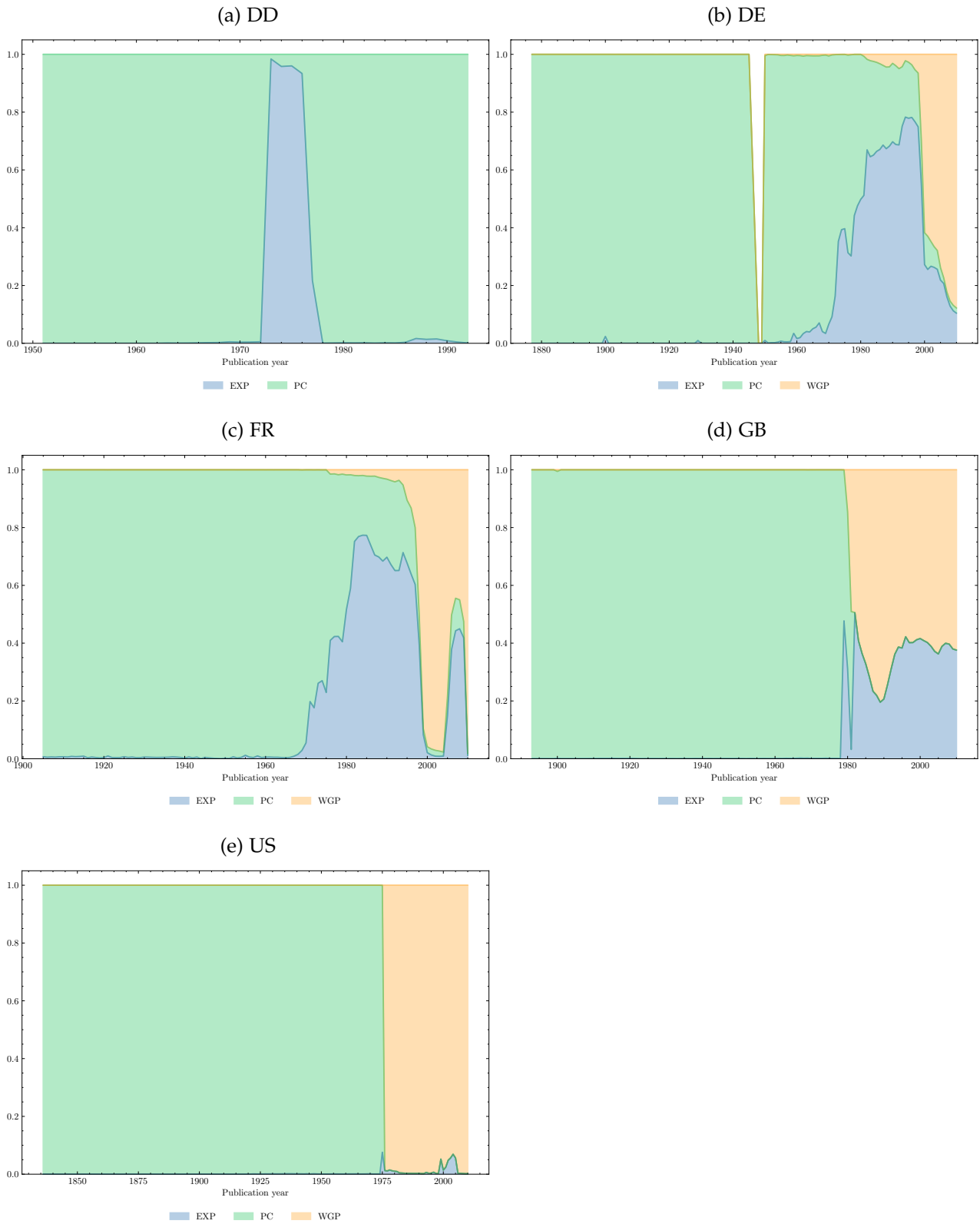


Figure A9: DATABASE COMPOSITION BY SOURCE (NUMBER OF PATENTS)



Notes: PC refers to PatentCity data, WGP refers to [de Rassenfosse et al. \(2019\)](#) data and EXP refers to data collected from family expansion from patents included in either PC or WGP.

Figure A10: DATABASE COMPOSITION BY SOURCE (IN SHARE)



Notes: PC refers to PatentCity data, WGP refers to [de Rassenfosse et al. \(2019\)](#) data and EXP refers to data collected from family expansion from patents included in either PC or WGP.

Figure A11: DATABASE COVERAGE BY OFFICE AND PUBLICATION YEAR (IN ABSOLUTE VALUES)

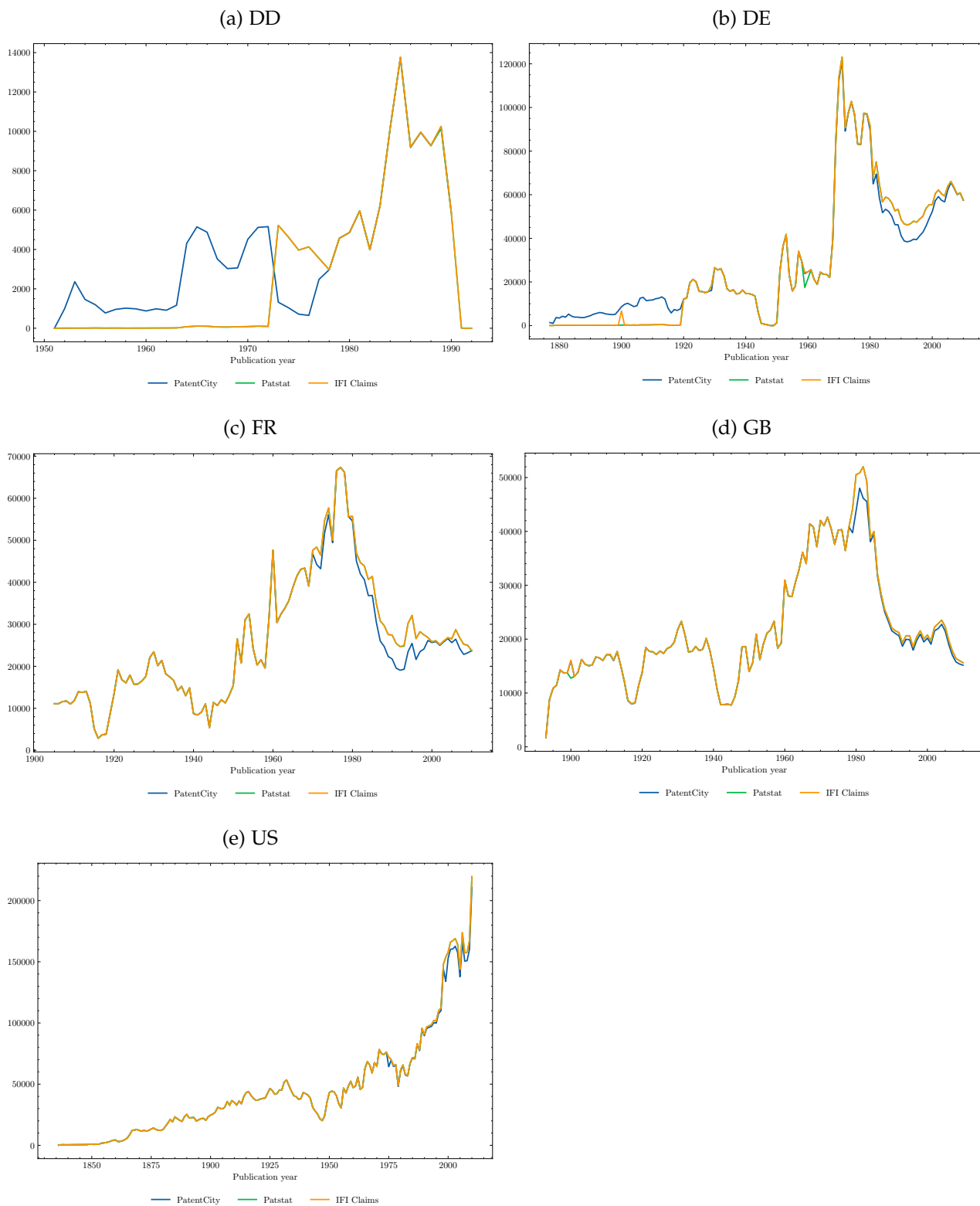
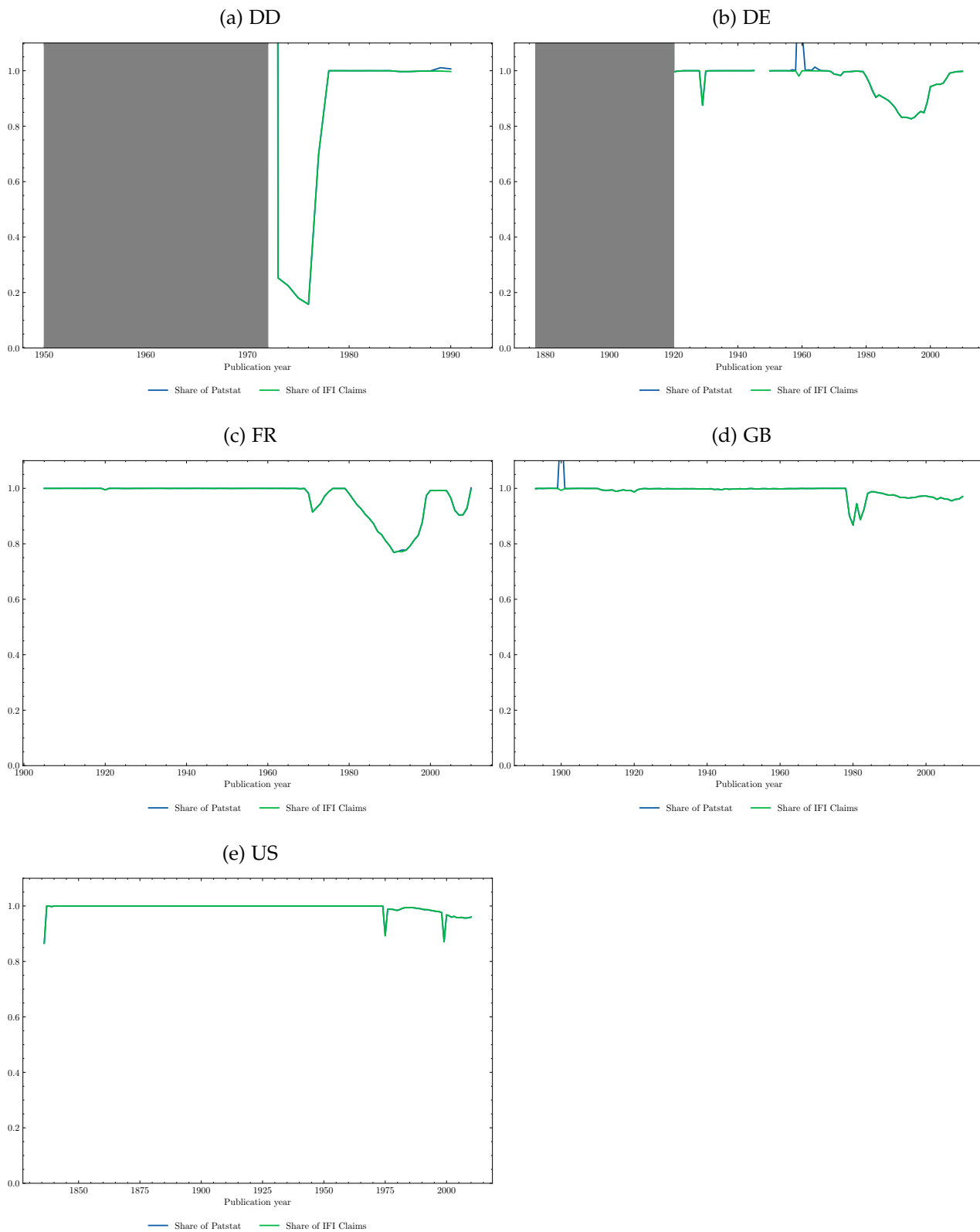


Figure A12: DATABASE COVERAGE BY OFFICE AND PUBLICATION YEAR (IN SHARE OF THE CLAIMS DATABASE COVERAGE)



Notes: We report the share of patents which are reported in our database at the office-publication year level as compared to the coverage of the IFI Claims database (publicly available as part of the Google patents public dataset). Shaded areas correspond to office and publication years where patents reported in the IFI Claims database miss dates, meaning that we miss a proper denominator.

A.5 Population data and geographical adjustments

We construct population data at the regional level using official sources for the most recent period from as long as possible: respectively De-Statis for Germany, the Insee for France, the ONS for the United Kingdom and the Census Bureau's Population Estimates for the US. We summarize the source and construction below.

United States We start from the Census' population estimate at the county level 1969-2020 which we backdate using the census estimates every ten years since 1790. We aggregate the data at the current county boundaries using the crosswalk provided by [Eckert et al. \(2020\)](#).

Germany We begin with the official estimates provided by De-Statis, which are available yearly starting from 1977. To estimate population figures for earlier years, we use Census estimates from 1871, 1900, 1905, 1910, 1925, 1939, 1950, 1961, 1964, and 1974. In order to account for changes in territorial boundaries, we group the 401 NUTS 3 regions into 288 units by merging the most populous city-regions (*Stadtkreise*) with the surrounding regions using the procedure described in [Wyrwich \(2020\)](#).

France [The Insee](#) provides yearly estimates of population at the city level as well as for the years 1876, 1881, 1886, 1891, 1896, 1901, 1906, 1911, 1921, 1926, 1931, 1936, 1954, 1962, 1968, 1975, 1982, 1990 and 1999. We aggregate these at the NUTS 3 (*département*) level. To account for changes in the Paris area, we merged current départements 91, 78 and 95 into one (which broadly correspond to the former *Seine-et-Oise* region before 1968) and current départements 92, 93, 94 with Paris to form the former *Seine* département.

United Kingdom [The ONS](#) provides yearly estimates of population by local authorities since 1981 which can be aggregated at the NUTS 3 level using the crosswalk provided in the [repository of the project](#). Unfortunately, we were not able to construct historical estimates at the NUTS 3 level due to the very high number of boundary changes and only report estimates aggregated at the NUTS 2 level (Northern Ireland, counties in England, groups of districts in Greater London, groups of unitary authorities in Wales, groups of council areas in Scotland). We use the census in years 1871, 1881, 1891, 1901, 1911, 1921, 1931, 1939, 1951, 1961, 1966 and 1971 which we retrieved from [Vision of Britain \(VoB\)](#) population data, except for London where we use data directly available from the [London Census](#) and for Northern Ireland where we used information from NISRA. Not all NUTS 2 regions were available through VoB and we proceed to the following assumptions:

- Tweeddale was constructed by aggregating Peebles and Selkirkshire
- Roxburgh Ettrick and Lauderdale was constructed using Roxburghshire + Selkirkshire + one fourth of Berwickshire and one fourth of Midlothian
- Cheshire was obtained by aggregating Halton, Warrington, Cheshire east, Cheshire West and Chester

- Mid Glamorgan was constructed by taking half of Caerphilly to which we added Bridgend, Merthyr Tydfil and Rhondda Cynon Taff
- South Glamorgan was constructed as the sum of Vale of Glamorgan and Cardiff
- Clwyd was taken as the aggregation of Flintshire, Wrexham and Denbighshire
- Dyfed was constructed using Carmarthenshire, Ceredigion and Pembrokeshire
- Gwent was constructed from Blaenau Gwent, half of Caerphilly, Monmouthshire, Newport and Torfaen
- Vale of Glamorgan was assimilated to Glamorganshire

A.6 Additional annotation guidelines

Tables [A3](#) and [A4](#) present additional representative examples of the rules we used to label the patents. See Section [2](#) and [the detailed guidelines](#) for [East Germany](#), [Germany](#), [France](#), [the United Kingdom](#) and [the United States](#).

A.7 Structure of the dataset

The dataset is publicly available both as a csv file and in SQL. The unit of observation is the patent, identifiable from the DOCDB publication number. Each patent is associated with a set of patentees (inventors or assignees) which have nested attributes: name, citizenship, location and occupation. The structure of the dataset is presented in Table [A5](#).

Table A3: ENTITY ANNOTATION GUIDELINES

Patent office	Entity	Content	Example
DD	ASG	Assignee full name	Inhaber: Rhône Poulenc S.A , Paris (Frankreich).
	INV	Inventor full name (<i>Erfinder</i>)	Erfinder: Dr. Karl Jellinek , WD
	LOC	Location of the assignee/inventor	Erfinder: Jean Auguste Phelisse, Lyon (Frankreich).
	OCC	Occupation of the assignee/inventor (academic title)	Dr. Elisabeth Kob, WD.
DE	ASG	Assignee full name	ANTON KLEBER in SAARBRUCKEN
	INV	Inventor full name (<i>Erfinder</i>)	Frutz Doring , Berlin-Frohnau ist als Erfinder genannt worden
	LOC	Location of the assignee/inventor	Demag Akt-Ges. in Duisburg.
	OCC	Occupation of the assignee/inventor (academic title)	Dipl-Ing Georg Werner Gaze, Ingolstadt
	CLAS	Technological class (German system)	KLASSE 49h GRUPPE 27 D 16736VI/49h
FR	ASG	Assignee full name	M. Robert John Jocelyn SWAN résidant en Angleterre
	INV	Inventor full name	(Demande de brevet déposée aux Etats-Unis d'Amérique au nom de M. Ladislas Charles MATSCH)
	LOC	Location of the assignee/inventor	M. Louis LEGRAND résidant en France.
	CLAS	Technological class (French system)	XII Instruments de précision 3 POIDS ET MESURES, INSTRUMENTS DE MATHEMATIQUES
GB	PERS	Person full name	Maxim Hanson Hersey , Lighting Engineer
	ORG	Firm full name	We, The Convex Incandescent Mantle Company Limited , Manufacturers
	CIT	The origin of the firm or citizenship of the person	a subject of the king of Great Britain and Ireland ,
	LOC	Location of the person/firm	Maxim Hanson Hersey, Lighting Engineer, of 145, Bethune Road, Amhurst Park, London N..
	OCC	Occupation of the person	Maxim Hanson Hersey, Lighting Engineer .
US	INV	Inventor full name	Be it known that I, JAMES M. GARDINER , ...
	ASG	Assignee full name	ASSIGNOR OF ONE-HALF TO SMITH FULMER
	LOC	Location of the assignee/inventor	residing at Mikkalo, in the county of Gilliam and State of Oregon
	CIT	Citizenship of inventor	JOHN SCHLATTER, a citizen of United States

Notes: Colored text corresponds to the entities that we seek to extract: red for inventors, purple for assignees, olive for locations, brown for citizenship and blue for occupation. An annotation guidelines available at <https://cverluise.github.io/patentcity/> (section Guides).

Table A4: RELATION ANNOTATION GUIDELINES

Patent office	Relation	Content	E.g
DD	LOCATION	Links an ASG/INV to a LOC	Rhône Poulenc S.A → LOCATION → Paris (Frankreich)
	OCCUPATION	Links an ASG/INV to an OCC	Dr ← OCCUPATION ← Elisabeth Kob
DE	LOCATION	Links an ASG/INV to a LOC	MARIUS ALBERT de DION → LOCATION → PUTEAUX (Seine, Frankr.)
	OCCUPATION	Links an ASG/INV to an OCC	Dr ← OCCUPATION ← KARL HENKEL
FR	LOCATION	Links an ASG/INV to a LOC	M.Frederic PERDRIZET → LOCATION → France (Gironde)
	CITIZENSHIP	Links an ORG/PERS to its CIT	Maxim Hanson Hersey → CITIZENSHIP → subject of the king of Great Britain and Ireland
GB	LOCATION	Links an ASG/INV to a LOC	Maxim Hanson Hersey → LOCATION → 145, Bethune Road, Amhurst Park, London N.
	OCCUPATION	Links an ASG/INV to an OCC	Maxim Hanson Hersey → OCCUPATION → Lighting Engineer
US	CITIZENSHIP	Links an INV/ASG to its CIT	WILLIAM H. BAKER → CITIZENSHIP → citizen of the United States
	LOCATION	Links an ASG/INV to a LOC	SEDWARD WILLIAM YOUNG → LOCATION → Tytherley, Wimborne, Dorset, England

Notes: Examples of relations between extracted entities for each patent office. Colored text corresponds to the entities extracted: red for personal inventors, purple for assignees, olive for locations, brown for citizenship and blue for occupations.

Table A5: DATABASE SCHEMA

Name	Description	Type	Nb non null
publication_number	Publication number.	STR	18,626,068
publication_date	Publication date (yyyymmdd).	INT	18,625,367
family_id	Family ID (DOCDB).	STR	18,625,353
country_code	Country code of the patent office.	STR	18,626,068
pubnum	Publication number.	STR	18,626,068
kind_code	Kind code.	STR	18,626,068
origin	Indicates the origin of the patentee data (PC: patentcity, WGP25: Worldwide Geocoding of Patent - slot 25, WGP45: Worldwide Geocoding of Patent - slot 45, EXP: expansion).	STR	18,626,068
patentee	Patentee	REC	18,626,068
__.is_inv	True if the patentee is an inventor, else False.	BOOL	45,537,241
__.is_asg	True if the patentee is an assignee, else False.	BOOL	45,537,241
__.name_text	Name.	STR	43,402,865
__.person_id	Person ID (PATSTAT).	INT	23,763,520
__.name_start	Name start.	INT	19,639,345
__.name_end	Name end.	INT	19,639,345
__.occ_text	Occupation text.	STR	1,354,930
__.occ_start	Occupation start.	INT	1,354,930
__.occ_end	Occupation end.	INT	1,354,930
__.cit_text	Citizenship text.	STR	3,996,958
__.cit_code	Citizenship code.	STR	3,861,775
__.cit_start	Citizenship start.	INT	3,996,958
__.cit_end	Citizenship end.	INT	3,996,958
__.loc_text	Location text.	STR	42,232,737
__.loc_start	Location start.	INT	16,334,841
__.loc_end	Location end.	INT	16,334,841
__.loc_addressLines	Formatted address lines built out of the parsed address components.	STR	16,003,816
__.loc_locationLabel	Assembled address value for displaying purposes.	STR	41,901,699
__.loc_country	ISO 3166-alpha-3 country code.	STR	41,898,330
__.loc_state	First subdivision level(s) below the country. Where commonly used, this is a state code (for instance, CA for California).	STR	41,428,298

Continued on next page

__.loc_county	Second subdivision level(s) below the country. Use of this field is optional if a second subdivision level is not available.	STR	34,200,971
__.loc_city	Locality of the address.	STR	40,391,684
__.loc_district	Subdivision level below the city. Use of this field is optional if a second subdivision level is not available.	STR	18,276,320
__.loc_subdistrict	Subdivision level below the district. Used only for India.	STR	16,003,816
__.loc_postalCode	Postal code.	STR	23,837,493
__.loc_street	Street name.	STR	18,145,660
__.loc_building	Building name.	STR	16,130,485
__.loc_houseNumber	House number.	STR	17,710,245
__.loc_longitude	Longitude.	FLOA	41,517,796
__.loc_latitude	Latitude.	FLOA	41,517,796
__.loc_relevance	Indicates the relevance of the results found; the higher the score the more relevant the alternative. The score is a normalized value between 0 and 1.	FLOA	12,203,353
__.loc_matchType	Quality of the location match. pointAddress: Location matches exactly as point address. interpolated: Location was interpolated.	STR	41,268,017
__.loc_matchCode	Code indicating how well the result matches the request. Enumeration [exact, ambiguous, upHierarchy, ambiguousUpHierarchy].	STR	16,003,816
__.loc_matchLevel	The most detailed address field that matched the input record.	STR	41,643,215
__.loc_matchQualityCountry	MatchQuality provides detailed information about the match quality of a result at attribute level. Match quality is a value between 0.0 and 1.0. 1.0 represents a 100% match. Here, matchQuality is defined at country level.	FLOA	2,658,311
__.loc_matchQualityState	Same at state level.	FLOA	6,553,671
__.loc_matchQualityCounty	Same at county level.	FLOA	1,547,347
__.loc_matchQualityCity	Same at city level.	FLOA	11,331,772
__.loc_matchQualityDistrict	Same at district level.	FLOA	1,361,402

Continued on next page

__loc_matchQualityPostalCode	Same at postalCode level.	FLOA	147,862
__loc_matchQualityStreet	Same at street level.	FLOA	2,452,802
__loc_matchQualityHouseNumber	Same at houseNumber level.	FLOA	1,034,844
__loc_matchQualityBuilding	Same at building level.	FLOA	410
__loc_key	Key used for statistical area mapping (internal use).	STR	31,137,221
__loc_statisticalArea1	Name of the high level Statistical Area.	STR	31,061,188
__loc_statisticalArea1Code	Code of the high level Statistical Area.	STR	31,061,188
__loc_statisticalArea2	Name of the mid level Statistical Area.	STR	31,061,165
__loc_statisticalArea2Code	Code of the mid level Statistical Area.	STR	19,738,673
__loc_statisticalArea3	Name of the low level Statistical Area.	STR	31,055,300
__loc_statisticalArea3Code	Code of the low level Statistical Area.	STR	31,067,057
__loc_recId	Identifier of the input address in the response.	STR	42,232,737
__loc_seqLength	Number of results for the corresponding input record.	INT	12,244,380
__loc_seqNumber	Consecutively numbers the different results for the corresponding input record starting with 1.	INT	29,657,332
__loc_source	Geocoding source (in [HERE, GMAPS, MANUAL]).	STR	41,901,712
__is_duplicate	True if a patentee with the 'same' name has been detected in the same patent. Only one of the two is marked as duplicate.	BOOL	3,985,815

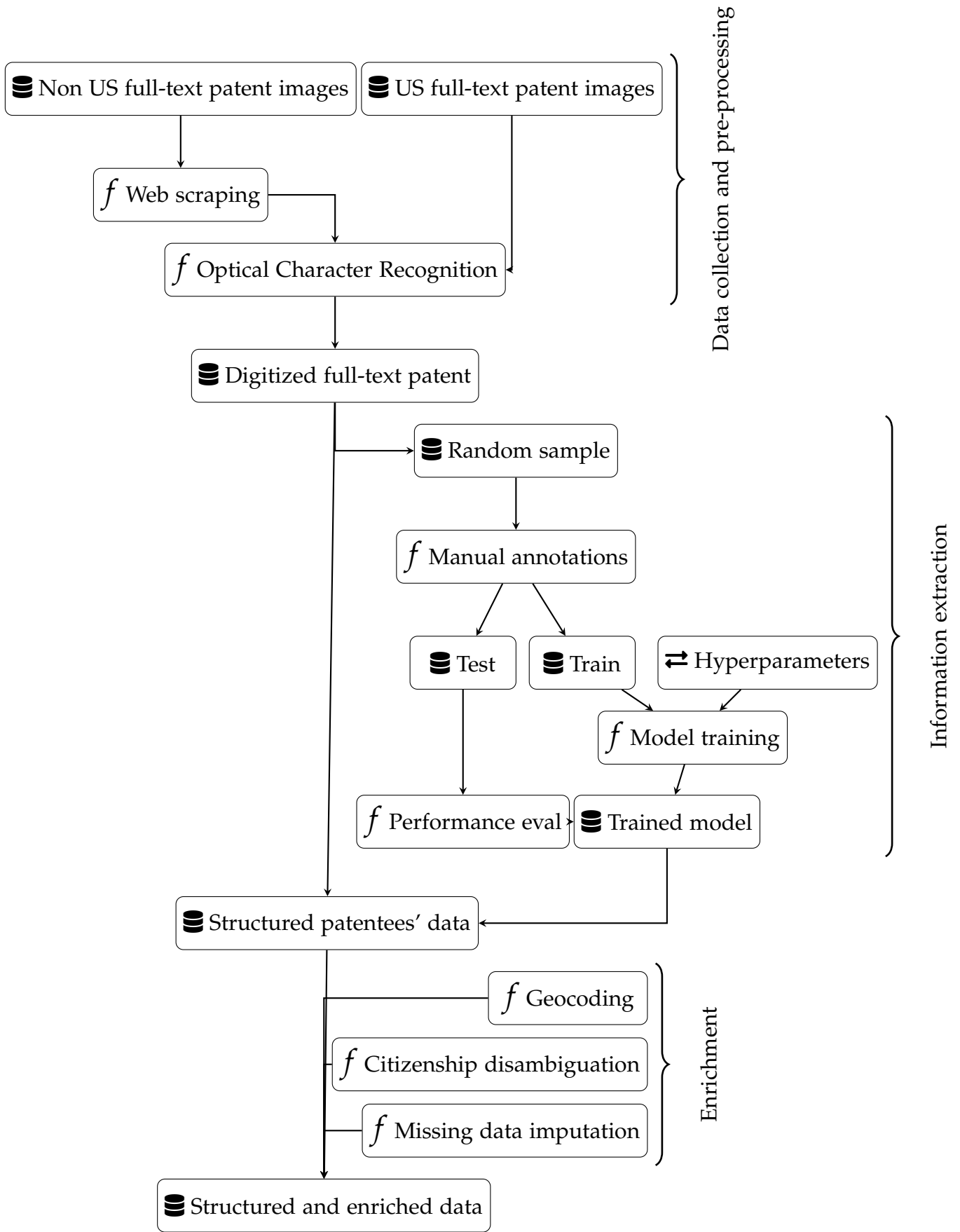
Notes: Variable names prefixed by a «__» are nested variables. For example, «__is_inv» is nested in the «patentee» variable.

A.8 Pipeline

We summarize the full pipeline from the raw documents to the structured and enriched database in Figure [A13](#).

A13.

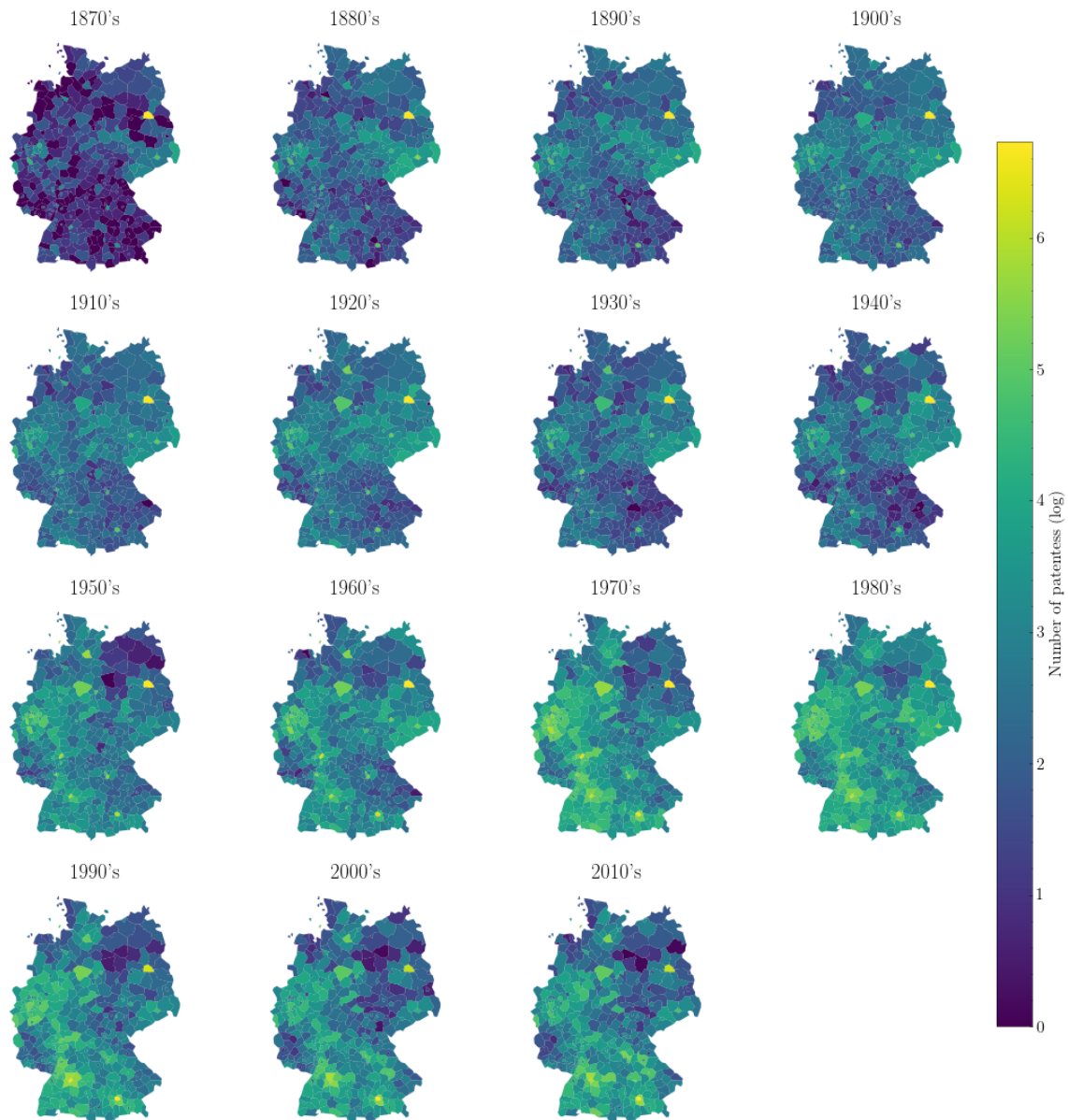
Figure A13: Workflow pipeline



B Additional Maps

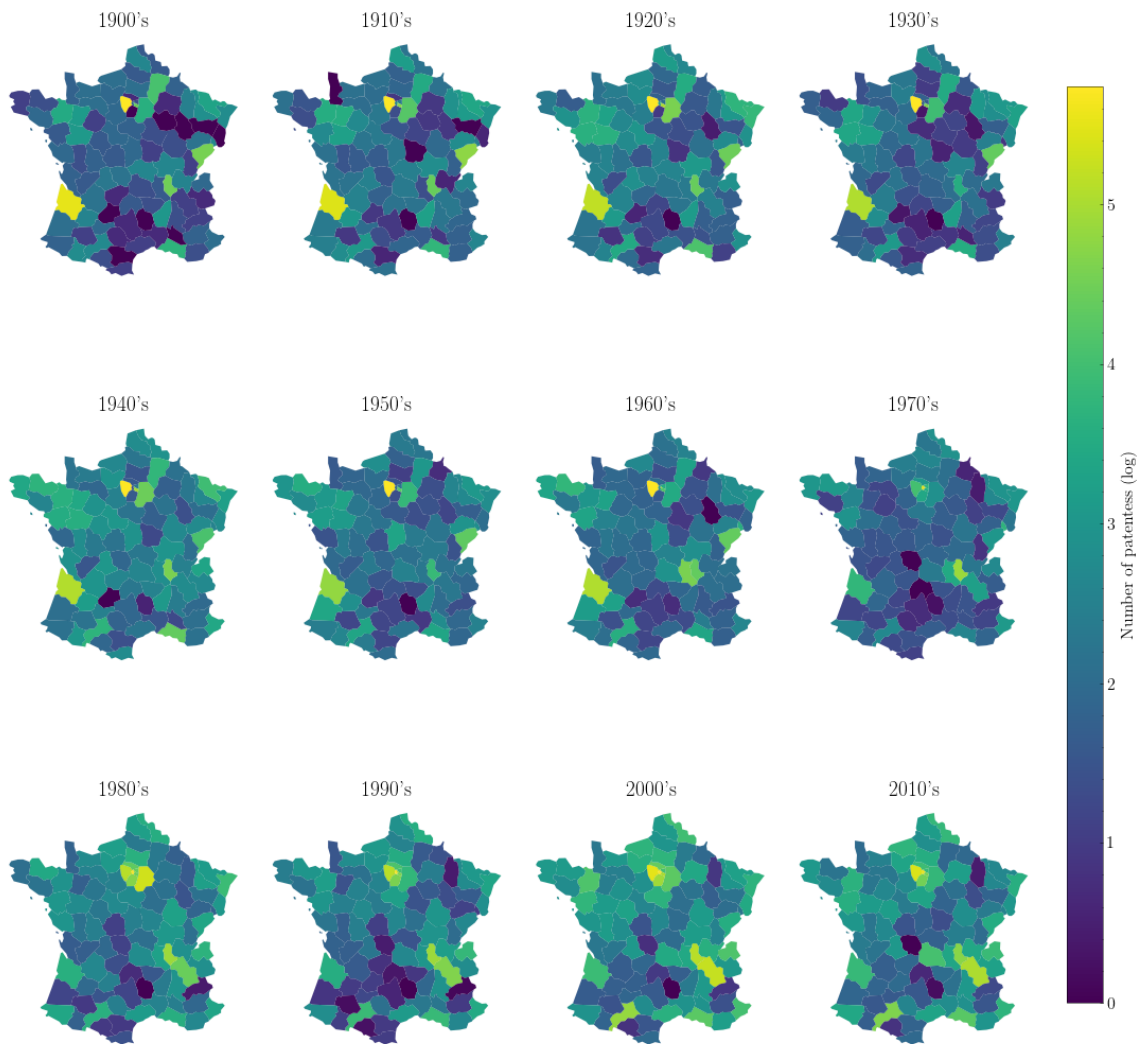
Figures B1, B2, B3 and B4 map the number of patentees by regions NUTS 3 (commuting zones in the US) by decade.

Figure B1: PATENTEES BY REGIONS AND DECADE - GERMANY



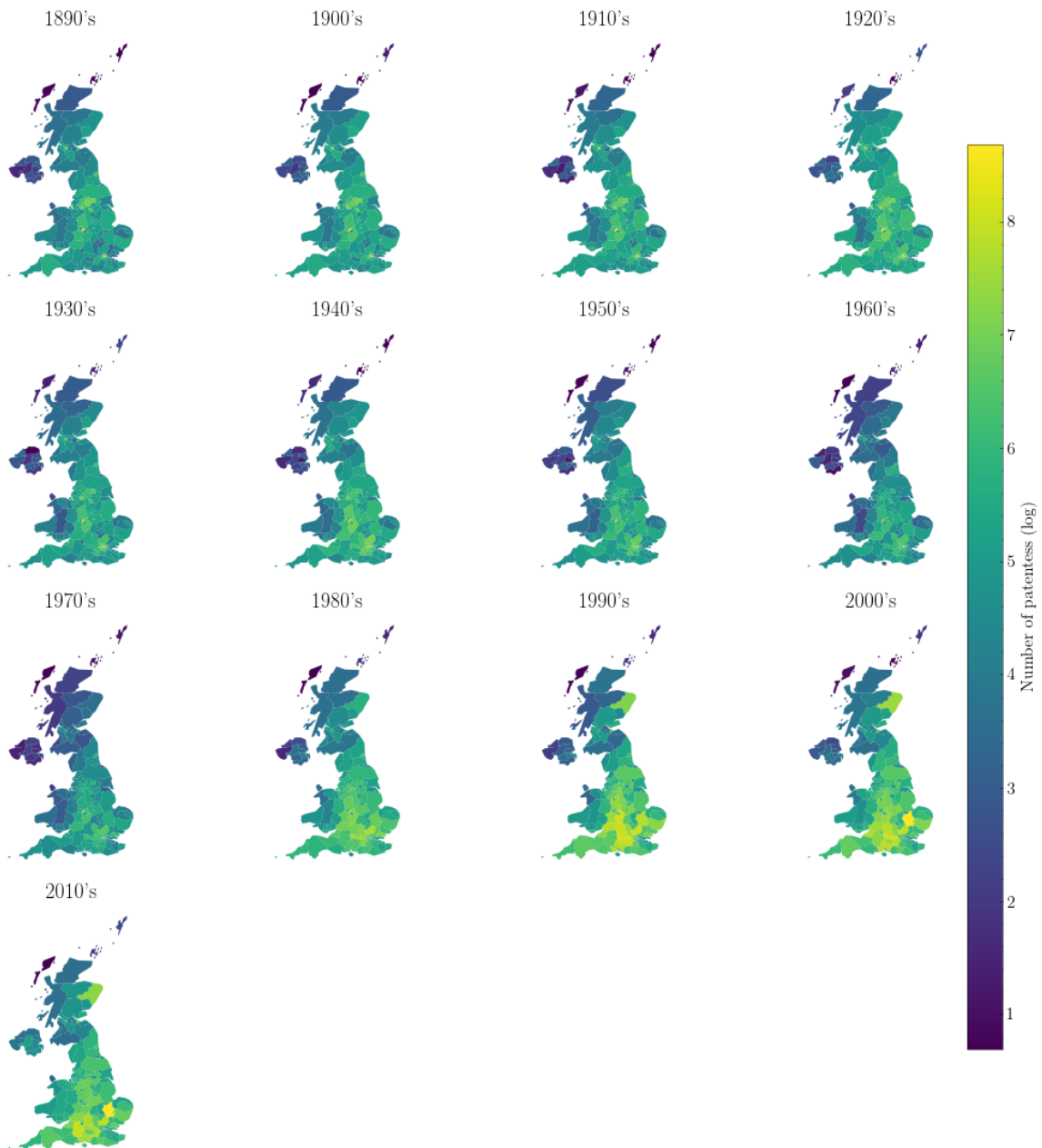
Notes: this Figure maps the total number of patentees (whether assignees or inventors), in log, for each county in Germany (*Kreise*) for each decade. The number of patentees is taken as a total over the full set of domestic patentees that are located at least at the county level.

Figure B2: PATENTEES BY REGIONS AND DECADE - FRANCE



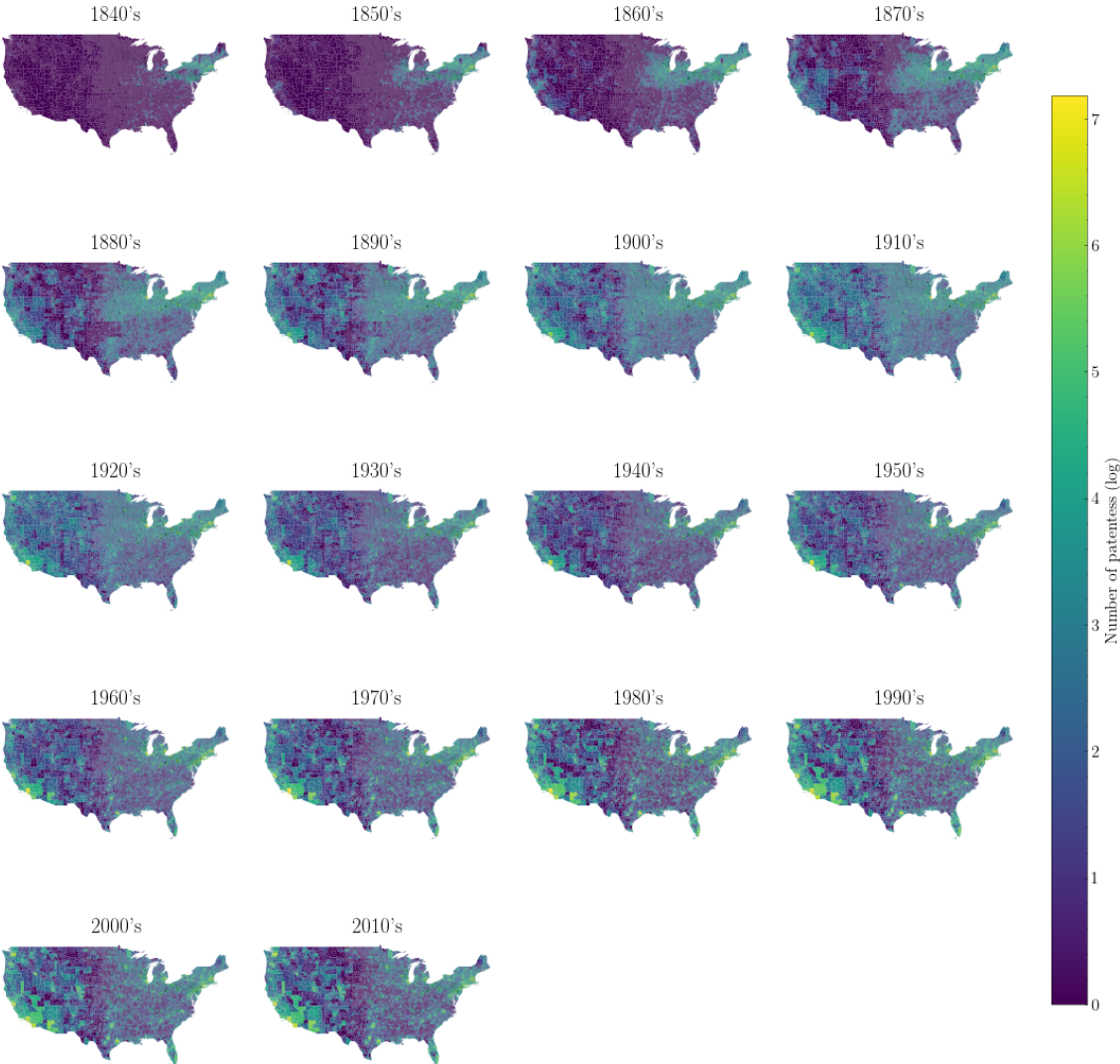
Notes: this Figure maps the total number of patentees (whether assignees or inventors), in log, for each county in France (*département*) for each decade. The number of patentees is taken as a total over the full set of domestic patentees that are located at least at the county level.

Figure B3: PATENTEES BY REGIONS AND DECADE - UNITED KINGDOM



Notes: this Figure maps the total number of patentees (whether assignees or inventors), in log, for each county in the UK (NUTS 3 regions) for each decade. The number of patentees is taken as a total over the full set of domestic patentees that are located at least at the county level.

Figure B4: PATENTEES BY REGIONS AND DECADE - UNITED STATES



Notes: this Figure maps the total number of patentees (whether assignees or inventors), in log, for each county in the USA for each decade. The number of patentees is taken as a total over the full set of domestic patentees that are located at least at the county level.

C Brief history of the patent systems

This Appendix briefly summarizes the major changes in the patent systems of the five patent offices considered in this article, focusing on the period ranging from the 19th century to 1980. Much more details can be found in [Khan and Sokoloff \(2001\)](#); [Van Dulken \(1999\)](#); [Khan and Sokoloff \(2004\)](#); [Moser \(2005\)](#); [Bert \(1960\)](#); [Hipp et al. \(2022\)](#); [Davenport \(1979\)](#); [Dobyns \(2016\)](#) and reference therein.

United States

- July 1790: The USPTO granted its first patent. A patent is defined as “any useful art, manufacture, engine, machine, or device or any improvement thereon not before known or used”. The fee for obtaining a patent was set at \$3.70 and its terms was set to 14 years (with a possibility of a 7 year extension).
- 1793: Significant changes were made to the patent law. Applications were no longer examined for novelty but were granted as long as the fees were paid. The fees were raised to about \$30 to which some small additional costs should be added for each page of the patent description.
- 1802: The patent office became an independent bureau within the Department of State.
- July 1836: The examination system and novelty criteria were reestablished. A search process for prior art was implemented, and appeals were permitted. The fees were higher for foreigners, with British citizens paying \$500 and others paying \$300. This is generally view as the establishment of the modern patent system in the US.
- December 1836: A fire destroyed all patent records and drawings. Around 2,800 of these have been recovered, mainly from the inventors’ copy. These are now known as the X-patents. Fees were about \$30.
- 1861: The term of a patent grant was extended from 14 to 17 years (and stayed at 17 until 1994). The fees for obtaining a patent were raised to \$35 (\$15 at the time of application and \$20 when granted). All discrimination against foreign applicants are eliminated, except for countries discriminated against the United States
- 1870: The 1870 Patent act consolidated the previous acts and strengthened the authority of the patent office. The USPTO started to register trademarks
- 1887: The United States becomes a member of the Paris Convention which guaranteed the protection of US inventions abroad and conversely
- 1921: To help European countries following WWI, the Nolan Act provided foreign inventors some benefits regarding time of application and fees
- 1922: The fees for filling a patent increased from \$15 to \$20. The total fees were about \$40 dollars.

- 1930: Application fees are raised to \$25 and remain relatively stable in real terms until the end of the 1960s (De Rassenfosse and van Pottelsberghe de la Potterie, 2013)
- 1952: Modernization of the patent law. Inventors were required to not only describe their invention but also the grounds for its infringement. In addition, for a patent to be granted, the invention had to be novel, useful, and non-obvious.
- 1967: Fees increase from 60 to 239 dollars
- 1968: The Patent Cooperation Treaty (PCT) is signed. This treaty provides a unique procedure to file a patent application in all member states.
- 1980: The Patent and Trademark Law Amendments Act, also known as the Bayh-Dole Act was enacted and authorizes contractors to retain ownership of inventions resulting from research funded by the federal government.

France

- 1791: The first patent legislation was enacted in France. Applications can be filed through a registration system without any examination. The inventor could chose a period of 5, 10 or 15 years. The price of a patent ranged from 300 to 1500 francs.
- 1844: The initial act was amended in 1800 and 1844. The fees remained in the same order of magnitude as before: 500 francs for a 5 year protection and 500 additional francs of each additional 5 years up to 15. The fees were payable annually. The law of 1844 changed the nature of the patent right. The patent is granted on the sole basis of the applicant's declarations and under his sole responsibility but could be voided if the patentee failed to implement its innovation within 2 years. Scientific discoveries or theoretical concepts can be patented under the condition that there is a clear industrial application.
- 1883: The Paris convention is signed. Foreign inventors enjoyed a 12 month priority period after their domestic application to apply for their patent in France. The country commit to create a centralized service of intellectual property and to edit a list of granted patents.
- 1901: The national patent office: "Office National de la Propriété Intellectuelle" (ONPI) is created. Until this year, the search for prior arts by inventors was very difficult as the specifications of patents were solely accessible in handwritten form in the office where they were initially deposited. The ONPI centralized all the journals and publications.
- 1902: All the patents are published *in extenso* which are typed and no longer hand-written and sold at the unique cost of 1 franc. Inventors could chose to delay the publication of their patent for up to one year.
- 1910: The total fees are still around 100 francs per year and paid annually
- 1919: The maximum term of a patent is set to 20 years

- 1951: The ONPI is replaced by the “Institut Nationale de la Propriété Industrielle” (INPI) , entirely self-financed and placed under the supervision of the Ministry of the Economy and Finance
- 1953: The legal obligation to implement one’s patent within 2 years is removed
- 1968: Modernization of the intellectual property law with the implementation of an examination system for all applications. The application is made public 18 months after its filing.
- 1978: Alignment of the French patent legislation with the European legislation following the creation of the European Patent Convention.

Germany

- 1877: Creation of the Imperial Patent Office (*Kaiserliches Patentamt*) in Berlin. Before this date each of all the 25 small states that came together to form the nation of Germany in 1871 was responsible for its own intellectual property law. The cost of a patent was high in order to filter trivial inventions: 30 marks during the initial year, 50 and 100 marks for the next two years and 50 additional marks per years for up to 15 years (in 1871 the average wage in the manufacturing sector was about 10 marks per week). German patent system included a mandatory examination by an official patent examiner. This is the first patent system in the world to do so.
- 1891: The Patent Act of 1891 enhanced the efficiency of the patent office through changes in its organization including the separation of the application division and appeal division, as well as in the administrative procedure. As a part of these changes, this Act introduced a “preliminary examiner” who was tasked with addressing any issues in an application before it went to the application stage. Utility models (*gebrauchsmuster*) are introduced promoted by the chemical industry which allows inventors to protect their products and not only the process that produced it. Utility models are viewed as a weaker version of examined patents and could be obtained through a registration system.
- 1903: The German Empire acceded to the Paris Convention
- 1919: The Imperial Patent Office is replaced by the *Reichspatentamt* (Patent Office of the German Reich).
- 1923: Maximum term is extended from 15 to 18 years.
- 1936: The Patent Act of 1936 reorganized the role of the patent office. Patent litigation and cancellation requests were centralized in the patent office and no longer in ordinary courts. In order to help small inexperienced inventor, the patent act introduced a “grace period” of six months which allows an inventor to fill for a patent even if its invention is already known to the public from less than 6 months. The Act also introduced the principle that the right to an invention belongs to the inventor
- 1945: The patent office is destroyed and remained inactive until 1950.

- 1949: The Office for Inventions and Patents (*Amt für Erfindungs- und Patentwesen*) opened in Soviet occupied Berlin. Meanwhile the Federal German Republic patent office moved from its destroyed Berlin's premises to Munich. The regulations concerning the protection of intellectual property rights were very similar in both of the German states but the main difference rest on the exploitation of the patent. In West Germany, the patent holder enjoyed full rights to exploit their patent. Conversely in East Germany, most patents were in fact economic patents (*Wirtschaftspatente*) and state-owned enterprises were authorized to use them. The inventor was only recognized as the intellectual source and remunerated in case the patent proved useful. Only a small shares of patents were provided exclusive rights to the applicant (*Ausschließungspatente*).
- 1968: Implementation of a publication of the new applications 18 months after the filing date.
- 1970: The Patent Cooperation Treaty is signed and implemented.
- 1977: The European Patent Convention came into effect
- 1980: The new Patent Act modernized and harmonized the intellectual property laws in Germany.
- 1990: Following the unification of Germany, the East German patent office was merged to the DPMA. Former Federal German Republic patents validity was extended to the whole German territory while former East German patents continued to be only effective in the former East German territories.

United Kingdom

- 1449: The first English "patent" was granted in 1449. During the 15th century, exclusive rights for trades and manufacturers, such as patents for inventions, were bestowed by The Crown through the grant of monopolies.
- 1624: England's first patent law, the Statute of Monopolies was established. The patent system then developed progressively. The term of a patent is 14 years.
- 1852: Modern patent law was established. The separate patent systems of Scotland and Ireland were abolished and a office was set in Chancery Lane. The Patent Office mandated a written account of the invention to be submitted and implemented the practice of publication of patent documents. Until this time, the cost of a patent was about £100 (and up to £350 for extended coverage to Scotland and Ireland) which is view as very high (almost the yearly wage of a skilled worker). This high cost was accompanied with very lengthy administrative procedures which reflected the general distrust of monopolies that prevailed in England (MacLeod et al., 2003). As a result, only 14,359 patents were granted between 1617 and 1852 (Van Dulken, 1999). The 1852 patent law reduced the cost of a patent dramatically to £25 but subsequent fees have to be incurred to renew the patent: £50 after 3 years and £100 after 7 years up to a maximum term of fourteen years. The cost of a patent in Britain was still considerably higher than in the US or in other European countries.

- 1883: The Patents, Designs and Trademark bill introduced a new simplification in the patent application procedure (for example, applications could be mailed rather than presented in person) and a further reduction in the cost of patents to £4 (£1 at the time of application and £3 at the time of granting). Additionally, the first renewal payment of £50 was postponed by one year (after 4 years). An option to smooth the renewal fees and pay every year was also introduced.
- 1884: The United Kingdom signed the Paris Convention
- 1902: Official examination for prior art by an examiner became mandatory before granting a patent. An additional £1 was charged for this service. The fees gradually increased to reach £85 in 1977. Total renewal fee costs also increased gradually to £824 in 1977.
- 1907: Patents applications can be rejected on the basis that they are considered “frivolous” or not novel enough.
- 1919: the term of a patent is extended to 16 years
- 1932: the rules under which a patent may be revoked are clarified and a patent appeal court is established.
- After World War Two there were various efforts to internationalize the patent system and increasing cooperation. The Patent Cooperation Treaty was signed in 1970 and the European Patent Convention in 1977 which opened the way to EPO patents.
- 1977: the Patent Act of 1977 is viewed as one of the main changes in the British patent system. It established a 20 year term, a two-stage review process which implies a publication after a 18-month period, in line with other European patent offices. The requirements in terms of novelty were raised to match international standards.

D Model Cards

Details on the performance of the model are given in the website of the project. Specifically:

- Model cards for [DD](#)
- Model cards for [DE](#)
- Model cards for [FR](#)
- Model cards for [GB](#)
- Model cards for [US](#)

Model performance are also summarized in Tables [D1](#), [D2](#), [D3](#), [D4](#) and [D5](#), respectively for East Germany, Germany, France, the United Kingdom and the United States.

Table D1: MODELS' PERFORMANCE BY FORMAT IN DD

Format	Metric	ALL	ASG	INV	LOC	OCC
1	p	0.99	0.99	0.96	0.99	0.99
	r	0.99	0.99	0.96	0.99	1
	f	0.99	0.99	0.96	0.99	0.99
2	p	0.95	0.94	0.95	0.98	0.94
	r	0.94	0.87	0.97	0.95	0.94
	f	0.95	0.91	0.96	0.96	0.94

Notes: Reported performance metrics were computed on the test set - unseen during training. The "Format" column indicates the different models used for the office. For the German office, there was a major shift in the patent information display in 1881 forcing us to train two different models. Performance metrics are reported as follows: precision/recall/F1-score.

Table D2: MODELS' PERFORMANCE BY FORMAT IN DE

Format	Metric	ALL	ASG	CLAS	INV	LOC	OCC
1	p	0.99	0.98	0.99	0.99	1	0.97
	r	0.99	0.99	1	0.96	1	0.98
	f	0.99	0.98	1	0.98	1	0.97
2	p	0.99	0.99	0.99	0.98	0.99	0.97
	r	0.98	0.98	1	0.99	0.98	0.97
	f	0.98	0.98	0.99	0.99	0.98	0.97

Notes: Reported performance metrics were computed on the test set - unseen during training. The "Format" column indicates the different models used for the office. Performance metrics are reported as follows: precision/recall/F1-score.

Table D3: MODELS' PERFORMANCE BY FORMAT IN FR

Format	Metric	ALL	ASG	CLAS	INV	LOC
1	p	0.97	0.99	0.93	0.99	0.99
	r	0.97	0.99	0.93	1	0.99
	f	0.97	0.99	0.93	0.99	0.99
2	p	0.98	0.98	-	0.99	0.99
	r	0.98	0.98	-	0.98	0.99
	f	0.98	0.98	-	0.98	0.99

Notes: Reported performance metrics were computed on the test set - unseen during training. The "Format" column indicates the different models used for the office. Performance metrics are reported as follows: precision/recall/F1-score.

Table D4: MODELS' PERFORMANCE BY FORMAT IN GB

Format	Metric	ALL	ASG	CIT	INV	LOC	OCC
1	p	0.93	0.93	0.96	0.95	0.92	0.9
	r	0.94	0.92	0.96	0.96	0.92	0.86
	f	0.94	0.93	0.96	0.96	0.92	0.88

Notes: Reported performance metrics were computed on the test set - unseen during training. For GB, only one model is used. Performance metrics are reported as follows: precision/recall/F1-score.

Table D5: MODELS' PERFORMANCE BY FORMAT IN US

Format	Metric	ALL	ASG	CIT	INV	LOC
1	p	0.98	0.94	0.98	1	0.98
	r	0.99	0.96	0.98	0.99	0.99
	f	0.99	0.95	0.98	0.99	0.99
2	p	0.98	0.96	0.98	1	0.98
	r	0.99	0.96	0.97	1	0.99
	f	0.98	0.96	0.98	1	0.99
3	p	0.97	0.96	0.97	0.99	0.97
	r	0.97	0.96	0.97	0.98	0.98
	f	0.97	0.96	0.97	0.98	0.98
4	p	0.99	0.99	-	1	0.99
	r	0.99	0.98	-	1	0.99
	f	0.99	0.98	-	1	0.99

Notes: Reported performance metrics were computed on the test set - unseen during training. The "Format" column indicates the different models used for the office. Performance metrics are reported as follows: precision/recall/F1-score.