

**EXPLORING TRADE GLOBALIZATION IN THE LONG RUN:
THE RICARDO PROJECT**

BÉATRICE DEDINGER & PAUL GIRARD
(Sciences Po, Paris)

Abstract. Initiated in 2004, the RICardo project focuses on bilateral trade data of all the world's countries from the early nineteenth century to the eve of the Second World War. The project includes the construction of a database and the creation of a website. It is a pioneering work in the field of historical trade statistics that aims at providing easy-to-access research material to the scientific community and at attracting a wider public to the history of trade relationships. The paper emphasizes the originality of our project and sums up its various aspects by reviewing previous trade databases, clarifying the main features of historical trade statistics, and describing the construction and use of the RICardo website.

Key words: trade globalization; history of international trade; historical trade statistics; trade databases; exploratory data analysis.

Many researchers, engineers, and (formerly) students have contributed to the completion of this project: Olivier Accominotti, Jung-Hyun Ahn, Lars-Fredrik Andersson, Matteo Azzi, Thomas Barré, Guillaume Daudin, Jérôme Destombes, Marc Flandreau, Clemens Jobst, Riitta Hjerppe, Odile Mazilu, Karine Onfroy, Raul Sanchez, Vera Quina, Riad Rezzik, Grégory Tible, Diego Zamuner, Mengying Du. The RICardo project has benefited from financial support provided by the ANR (Agence Nationale de la Recherche) and Sciences Po. We thank Jean-Pierre Dormois and Guillaume Daudin for their help and comments regarding this paper.

RICardo (RIC for *Research on International Commerce*) is a database that documents bilateral trade flows the world over covering the period 1800/30 to 1938. When the idea emerged, in 2004, it was limited to the collection of data every ten years for the needs of our research team, and when the project was officially launched in 2007, it was intended to be completed within four years. More than ten years after its birth, the project has taken an enlarged dimension by providing annual series of historical trade statistics on an open-access website.¹ The aim of this paper is to present the different stages of this undertaking from the collection of data to the web visualizations, by insisting upon the difficulties we have experienced, the choices made, and the solutions that have been developed.

The original concept for the RICardo project started from the realization that all existing historical trade databases suffer from various limitations. Either they provide values for countries' total trade only, or cover the period after 1870, or are limited to a selection of countries or regions (mainly Europe and North America). Three main hurdles may have deterred scholars from building a large historical trade database: the relatively apparent low benefits of work, the unreliability of trade data, and the unavailability of sources. The first point is far from negligible as already noted

by (Lewis 1981, 33)² and confirmed by our experience. Secondly, the warning of (Morgenstern 1963, 180) concerning the reliability of trade data discourages further from embarking on such an undertaking: “*Writers on all phases of foreign trade will have to assume the burden of proof that the figures on commodity movements are good enough to warrant the manipulation and the reasoning to which they are customarily subject*”. This argument will be addressed more thoroughly in a later section. The last obstacle pertains to the availability of sources. It becomes more difficult to unearth bilateral trade statistics as we go back earlier in time, especially before 1870. The starting point of the RICardo project was the discovery of a significant and neglected compilation of bilateral foreign trade data, the French *Extraits d’Avis Divers* published between 1829 and 1839. Its exhumation prompted a search for all extant publications of commercial statistics the world over that suggested a great part of archival material had not yet been incorporated into available compilations of world trade statistics. This motivated our request for financing from the French *Agence Nationale de la Recherche* to build a large bilateral trade database that would allow a renewing of research on trade globalization by covering the entire period from the early trade globalization era to the eve of the Second World War.

After data collection, our work focused on the construction of a relational database structure to aggregate original data into a standardized information system. In fact, data transcribed from a great variety of sources are very heterogeneous. For two reasons: the international scope of the database includes many different currencies that have to be converted into a common value reference to allow comparison; the numerous names used to describe the trade partners necessitate some normalization work. Finally, we have created an Exploratory Data Analysis web application to help the research community use data for quantitative analysis.

The paper is organized as follows. The first section compiles an inventory of existing historical trade databases to point out the innovative character of the RICardo project. The second section examines the general features of the raw material we are working with, namely trade statistics, and the sources they are extracted from. In the third section we go into the construction of the database and the exploratory tools that are provided on the RICardo website.

The innovative nature of the RICardo project

Trade statistics of a country can be viewed in different ways, more or less disaggregated: total trade (sum of exports and/or imports), bilateral trade (distribution of total trade by partner countries), trade by products (distribution of total trade by products), bilateral trade by products (distribution of trade with each partner country by products). The RICardo database combines two sets of data: total trade data and bilateral trade data. The project started with the era of database management – with software such as Access or Filemaker. As years went by, we considered the new technological opportunities of web applications and decided to extend the project to the creation of a datascape (Latour et al. 2012). The team was reorganized to include developers and designers, and the database was no longer seen as an end product but as an evolutionary base that will be open to new data. Therefore, the RICardo database cannot, strictly speaking, be compared to other trade databases as it is designed to evolve over time. The reader/user is invited to look at the overview of the scale and coverage of the base that will be regularly updated on the website. In this section, we would like to outline the innovative dimension of the RICardo project by reviewing the most authoritative historical trade databases to this day.

A short history of trade data compilations

Foreign trade statistics are among the oldest official statistics available and they have naturally attracted the attention of historians of trade relations and the international economy. Fiscal incentives explain the precocious interest of states in foreign trade. Furthermore, entry into another jurisdiction was more conspicuous and therefore easier to record in the case a transshipment was necessary, such as in harbours (Clark 1938, xi). But the systematic compilation and (later) publication of foreign trade statistics by state agencies was not organised before the end of the seventeenth century (Charles and Daudin 2015). In England, concomitant with the creation of the *Lords of Trade*, an *Office of the Inspector General of Imports and Exports* was set up in 1696 (Clark 1938, 3-4). The French government followed suit at the time of the peace of Utrecht (1713) and entrusted the *Bureau de la Balance du commerce* with the task of collecting data on imports and exports (Charles and Daudin 2011).

Statistical data collection started on a large scale after the Napoleonic wars. A growing number of European states set up statistical agencies, and interest in and taste for statistical science spread among publicists and policy makers; the first international congress of statistics was held in Brussels in 1853 and to 1876 nine took place in various European cities (Escaith 2015). Moreover, overviews and summaries on economic statistics started to be compiled, among others the dictionaries of (McCulloch 1844, 1851, 1866) and Mulhall (1884, 1886, 1892, 1899). On international trade more specifically, one of the most significant collection is the Neumann-Spallart series, *Übersichten der Weltwirtschaft*, published from the 1870s, which provides estimates of world trade from 1860 using official statistics, as well as journals, dictionaries and yearbooks. Several publications of this type appeared until the First World War, notably the *Statistical Abstract of Foreign Countries* published in 1909 by the Bureau of Statistics of the US Department of Commerce and Labor. This later summary was several years in the making and presents probably the first comprehensive survey of all available historical data on world trade (including bilateral trade flows during the first decade of the twentieth century). Also noticeable is the *Statistisches Jahrbuch für das Deutsche Reich* that, from the 1903 edition, published tables of total exports and imports of the world's countries starting 1883. After the First World War, the League of Nations centralised the compilation and publication of international statistics in its *Memoranda on the balance of payments and the balance of trade*. Since 1948, detailed bilateral and total trade databases are available from the International Monetary Fund's *Direction of Trade Statistics* and *International Financial Statistics*.

Projects dealing with the compilation of foreign trade data of all countries of the world have thus began to mobilize the attention of statisticians about two centuries ago. It is not until after WWII that scholars attempted the task of reconstituting long-term series of foreign trade for all the world's countries.

Databases on total trade (Maddison 1962)

Maddison's bulky article deserves notice as the first attempt since World War Two to estimate the value (at current and constant prices) of the world's and the main traders' total imports and exports from 1870 to 1960.³ Besides, it presents the advantage of reproducing annual series of foreign trade for the 'G14' of the time⁴ and for three regions (Western Europe, North America, Rest

of the World). The national statistical abstracts are prominent among the official publications used for compiling the series; corrections introduced by the author are discussed extensively (such as in the case of the Netherlands). In order to estimate world trade, Maddison relies on and revises the series established by (Lewis 1952) which are themselves based on an earlier compilation by (League of Nations 1945). For the period 1870-1913, Maddison converted national estimates of trade flows using the exchange rates provided by (Svennilson 1954) but he does not explain, however, how the 'Rest of the World' series were computed even though they add up with the two others to make up 'world trade.'

(Lewis 1981)

Some thirty years after his first, Lewis made a second attempt at estimating world exports over the period 1850-1913 with a view to observe time trends of world trade. As he himself remarked in the introduction, "*the paper has turned out differently from its original plan. It was to have been a long analytical exercise accompanied by a brief statistical appendix. Instead, preparing the statistical appendixes has taken all the time available; they have become almost a separate paper*" (Lewis 1981, 11).

His global results are relatively close to Maddison's, but his survey is more comprehensive as the country annual series he reconstructed are included in the database, i.e. the total exports at current and constant prices for 53 countries (20 in Europe, 15 in America, 10 in Asia, 6 in Africa, 2 in Australasia) and sources are systematically indicated. For the period 1880-1913 he consulted the aforementioned US *Statistical Abstract of Foreign Countries*, with additions from the British *Statistical Abstracts*, those of the League of Nations as well as Mitchell's *European Historical Statistics*. The author acknowledges the scarcity of information before 1880 and mentions the secondary sources he used such as the *Statesman's Yearbook* as well as McCulloch's *Dictionary*. He explains at length the deflation procedures he adopted and the extrapolation method for bridging the gaps in the original series, as well as discusses individual country cases. His results are compared with those published in Mulhall's *Dictionary* and League of Nations' *Industrialisation and Foreign Trade*. Figures are expressed in US dollars but the exchange rates used are not included.

(Mitchell 2007)

Prominent among the data series are the chapters on foreign trade in the *Historical Statistics* published by B. R. Mitchell, the first edition of which goes back to 1975 and which have been updated and revised regularly. They present the most complete coverage with data on total exports and imports for a large number of countries: 125 for the pre-1939 period (24 in Europe, 31 in the Americas, 39 in Africa, 24 in Asia and 7 in Australasia). Some of the series go back to the end of the eighteenth century and include each country's bilateral trade with its major partners (5 in general). In the introduction, the author lists the sources he used, consisting essentially in the national statistical abstracts. The main drawback of the Mitchell series 'for ready use' is the fact that they are expressed in national currencies and that the author does not provide any conversion tables.

Databases on bilateral trade

(Bairoch 1973, 1974)

For the purpose of observing the changing structure of European trade with the rest of the world since the onset of industrialisation, Bairoch reconstructed series of European total and bilateral

trade flows in current values over the 1800/30-1970 period. As his focus was on the geographical dispersion of European exports and imports, he aggregated country data at ten year intervals to build series by continent: Europe (Western and Continental), North America, South America, Asia, Africa, Oceania. When the sources provided data at constant prices (particularly for the pre-1850 period), these were converted to current prices using local index of exports prices or approximate information based on the movement of prices in other countries (Bairoch 1973, 30). In case of gaps, missing data were extrapolated on the basis of exports of countries of similar economic structure, or on future trends of the country's trade or, alternatively, on the undistributed residual world trade apportioned to the country's share in the total (Bairoch 1973, 31; Bairoch 1974, 599). The statistical appendix and the list of sources lack detail.⁵ The latter only mentions that each country's published trade statistics have been tapped, but that "*it would be too lengthy and would serve no purpose to list in extenso the sources used*" (Bairoch 1974, 607). It is beyond doubt that Bairoch deployed a considerable effort for his investigation and it is a pity that only a fraction of statistical compilation found way in the final version of his article; it could have proved useful to other scholars.⁶

(Barbieri, Keshk, Pollins 1996-2012)

The experiment closest to the present RICardo project was undertaken in the 1990s by Barbieri. Originally it aimed at collecting and aggregating all foreign trade data from 1816, but upon the realisation that "*trade data for the pre-1870 period are too scarce to make any meaningful analysis possible*" (Barbieri, Keshk, and Pollins 2008, 16-17), she reduced the scope of her enquiry considerably.

Originally, therefore, Barbieri's ambition was very close to ours. Upon closer inspection, however, it seems possible to improve substantially on the range and quality of the results presented therein. On the crucial period 1870-1913, the information provided seems limited to the quotations from the *Statesman's Yearbook*; on the interwar period, to the data published by the League of Nations; and after 1945 to those of the IMF. Furthermore, data are collected for sovereign states only, excluding trade of colonies. The country nomenclature has been adapted throughout the period to match that of the Correlates of War Project.⁷

The scope of the Barbieri database can be best described with some summary figures. On the period 1870-2009, the bilateral database (dyadic_trade_3.0) includes 791 491 entries, but 90% of all observations are for the 1940-2009 period. For the period 1870-1939, of the 78 626 recorded observations, only 20 186 entries are fully documented, the rest representing reconstructed flows. There are 43 reporting countries for 1870-1913 and 63 for 1920-39 while partner countries are 46 and 69 respectively.

RICardo vs recent works

Recent years have seen a tendency among researchers to build their own historical trade dataset for the needs of personal research. In those works, the collection of data primarily serves the purpose of answering a specific research question. (Mitchener and Weidenmier 2008) assess the effects of empire on trade from 1870 to 1913; (Jacks, Meissner, and Novy 2011) gauge the importance of bilateral trade costs in determining international trade flows by selecting six years from 1870 to 2000; (Pascali 2014) builds a dataset covering the period 1850-1900 to investigate the role of the adoption of the steamship in spurring trade after 1870; (Gowa and Hicks 2013, 2015) analyze the impact of war, institutions and politics on trade in the immediate pre-WWI and

interwar period; (Hugot 2015) explores the chronology and geographical pattern of the two globalizations from 1827 to 2012. The authors refer mainly to the British *Statistical Abstracts*, the Barbieri database, Mitchell's volumes, national statistical yearbooks, and the League of Nations yearbooks. Gowa and Hicks, Pascali, and Hugot have also relied on primary trade sources. The database of Hugot, the largest one, has 160 000 observations over 1827-1947. It should be noticed that it includes data from an earlier version of RICardo (Hugot 2015, 42). Gowa and Hicks have gathered about 50 000 data over the 1900-1938 period. The other databases contain between 20 000 and 30 000 data.

By quoting these works and previous databases, we would like to emphasize the original character of the RICardo project and the fact that it is not easily comparable with these scholar works. First, the primary aim of RICardo is to provide the researcher with a ready-to-use and comprehensive compilation of historical bilateral trade statistics so that he does not spend part of his time redoing part of the job made by others. Second, the size of the RICardo database is not fixed. Thanks to a versioning process presented in the next section, new data will be regularly added. Third, RICardo data are accessible through a website to a wide audience. Furthermore, RICardo will be open for collaboration with other trade database projects so that it can become a reference in the field of historical bilateral trade statistics. A first example of such cooperation will be experimented with the trade dataset of (Federico-Tena 2016). They have estimated the most exhaustive world exports and imports series since Lewis's attempt. The database they have built for the period 1800-1938 includes a total of 241 polities (Federico-Tena 2016, 8-9)⁸ and provides four types of data for each total import and export flow: at current prices and current borders; at current prices and constant borders (1913); at constant prices and current borders; at constant prices and constant borders. The series at current prices and current borders will be visible on the RICardo website in a comparative way (see below section 3).

General features of RICardo's data

Shortcomings of historical trade statistics

Trade statistics are government statistics that originate from customs declarations about the quantity and/or the value or unit price of the merchandises that enter and exit the customs territory (the customs territory does not necessarily correspond to the 'national' or political territory of the entity). The customs declarations are sent to a central statistical office whose function is to evaluate, classify and publish national trade statistics (Charles and Daudin 2015). The data contained in these publications are RICardo database's raw material. It is customary to warn prospective users of the pitfalls involved in the handling of foreign trade statistics, even more so when they go back in time, and even more if they include bilateral flows.⁹ (Federico and Tena 1991) revisited the question by focusing on the information available on total trade flows. The authors compare total exports and imports of a country relative to the total sum obtained from and sent to its partners over the same years. They reach the conclusion that external trade statistics were quite reliable and that this reliability improved during the interwar years thanks to the standardisation promoted by the League of Nations. But they also confirm the impression of many authors that "*greater caution should be taken when handling data on the geographical distribution of trade, which are usually rather unreliable.*" A more recent reconsideration of the problem added

an optimistic note to this statement regarding the reliability of statistics relating to bilateral trade of textile goods in 1913 (Marín 2005).

It is therefore necessary to be familiar with the content and make-up of historical trade data. But the object of this paper is not to present exhaustively all the characteristics of such data. We focus here on the nature of the sources of information and the essential features and limitations of the data.

Data collection and sources

The guideline we adopted when we began the project consisted in tapping ‘upstream’ publications by national trade supervising authorities. However, human and financial constraints led us to revise our top ambitions and to use alternatively “primary” (national customs returns) and “secondary” (compilations of primary) sources. Besides, we were inclined to think that secondary sources were more or less of the same quality as primary ones. Eleven people were mobilized to collect the first set of data. We gave priority to French libraries (mainly the *Bibliothèque nationale de France* and the *Institut national de la statistique et des études économiques*); we went to Berlin, London, Madrid, and Moscow to get complementary sources; and we benefited from the assistance of R. Hjerpe and L.-F. Andersson for Finnish and Swedish trade statistics. At the end of 2015 the total number of observations in the database amounted to 267 000 (one observation = one bilateral trade flow of exports or imports) of which about 20 000 were missing flows. An overview of the database then brought to light the most obvious flaw of our strategy. There was a significant decrease in the number of total observations over the interwar period mainly due to the use of the League of Nations sources. It proved that primary and secondary sources were not of similar quality. In fact, national trade tables generally offer a greater wealth of detail, especially with regard to bilateral flows. We therefore decided to return to our core ambition by progressively including more primary sources, either through our own efforts or through integration of external contributions. An updated overview of the complete set of sources can be seen on the RICardo website. We describe in more detail below the secondary sources.

Annales du commerce extérieur

France has been a major player in international trade in the nineteenth century and it should not come as a surprise that its role is reflected by a commensurate number of publications on this subject. These particular series include the reports written by French consuls in different trading posts around the globe. They have been tapped by historians but rarely been used for establishing series of data. Most foreign trade specialists are familiar with the 3rd series of “*Avis divers*” but it has been possible to trace the 1st and 2nd series, which give a quantitative overview of bilateral trade of France’s foreign partners as early as 1829. There are indications in the *Archives Nationales* that individual reports began to be consigned in 1815 and the first issue available there is dated July 1821.

The *Extraits d’Avis Divers, 1^e série des Avis divers* (1829-1839), include some forty volumes, intended primarily for Chambers of Commerce and government agencies. They consist of economic and trade reports written by French consuls in different trading posts around the globe. These reports are usually not easy to locate and it has not been possible to piece together the whole series to-date. It is of great interest to us as the authors have started to make systematic use of foreign trade statistics, including breakdowns by partner countries. They include all relevant

information collected by the Ministry of commerce on the development of industry, trade and navigation in France's foreign partners as well as on their trade and customs legislation.

The *Bulletin du Ministère de l'Agriculture et du Commerce, 2^e série des Avis divers* (1840-1842), is made up of three volumes each including a number of individual monographs. It was intended to serve as a reference on foreign tariffs and customs regulations of the various countries France entertained relations with and to provide interested parties with information on the economic situation and trends in commercial policy of these countries.

The *Annales du commerce extérieur, série Faits commerciaux, 3^e série des Avis divers* (1843-1883), is the most extensive, consisting of some 100 volumes, each dedicated to a particular political entity (52 have been identified in total over the whole period). They provided relevant information for French traders and officials on the economic and commercial position of the country as well as numerous and detailed statistics, especially regarding bilateral trade flows.

The *Annales du commerce extérieur, Commerce et navigation des principaux pays étrangers* (1873-1917) is the last series. It adopted the format of the British statistical abstracts and offered tabulated statistical summaries relative to the various countries of the world. The data therein have been obtained from the official customs records of these countries and expressed in French francs – conversion rates being given in each volume. It has been observed that these series offer a less comprehensive survey than the Statistical Abstracts which include a greater number of reporting and partner countries.

British Statistical Abstracts

These well-known series was the usual basis for the historical reconstruction of world trade accounts in the nineteenth century. Each volume includes total exports and imports by country for a relevant period as well as the geographical distribution of each country's foreign trade. We have extracted from this source all available data of bilateral flows except for those countries for which we had the official publication supplying this type of information. In certain instances, however, the *Abstracts* have proved useful to check and, when necessary, supplement a country's total trade figures.

The *Statistical Abstract for the several Colonial and other Possessions of the United Kingdom in each year from...*¹⁰ contain 70 volumes published from 1865 to 1950 and covering a period from 1850 to 1947, in which 92 entities are identified. Each volume presents a varying number of summary tables of countries' imports and exports, separating movements of goods and of precious metals. Other tables present a geographical breakdown of each country's foreign trade.

The *Statistical Abstract for the Principal and Other Foreign Countries in each year from...* contain 40 volumes published from 1874 to 1924 and covering the period 1860-1918, in which 101 state entities are identified. For each individual state, statistics of total trade are expressed in national currency as well as in pound sterling, the conversion table appearing in a separate appendix. Bilateral trade statistics appear from volume 3.

Memoranda on the balance of payments and the balance of trade¹¹

The economic committee of the League of Nations noted in a report dated December 1925 that “in the course of its day-to-day activities, many agencies of the League publish a vast quantity of statistics which are of considerable interest to governments and businessmen but that these compilations remain scattered in various publications and that there could be a great advantage in collecting the statistics thus produced in one single volume” (League of Nations 1927, 6). A number of monographs on various subjects were published at the beginning of the 1920s, among them a

memorandum on the foreign trade statistics of various member countries. This survey, conducted on an annual basis, was continued under various titles until 1938. As mentioned above, this was the source we tapped to document total as well as bilateral trade flows for the period spanning 1924 to 1938. However, League of Nations' trade data are incomplete. Not all the world's countries are represented (only 64 to 71 depending on the year) and the list of partner countries for each reporting country is also simplified. For example, the 1929 *Memorandum* of the League of Nations provides figures for 43 trading partners of the United Kingdom (89 % of total trade) whereas the *Annual Statement of Trade* records 165 partner entities. As a result, we decided to perfect the database by collecting new data from national trade statistics. This is a work in progress.

General or special trade

International trade flows have been discussed so far without specifying the content of the term. Being a statistical construction, the notion of a country's 'external' or 'foreign' trade may vary from period to period or country to country, which complicates tracking aggregate trade flows over the long-term (Maizels 1953, 44-50).

Generally speaking internationally traded goods brought into or out a so-called *customs territory* follow certain procedures, which register them into different trade types according to their next destination. Imported goods reaching their port of discharge either cross the border (after submitting to customs regulations) and are put at the disposal of the importer, or remain under customs control. In the former case, imported goods are intended for merchandising without any further processing being performed, are introduced into the domestic distribution network to be put on sale for final consumption, or are put into a production network towards further transformation (the new goods being intended either for home or foreign consumption). In the latter case, the goods can be stored in bonded warehouses, in which case they are either re-exported, introduced into the domestic distribution network (after payment of the pertaining customs duties), transformed for later exportation; or they enter the transit or transshipment to another foreign country. As regards merchandise exports, goods are classified into four possible categories: goods produced within the exporting country (including imported goods having undergone some degree of transformation); imported goods having not undergone any transformation (re-exports); goods stored in bonded warehouses exported with or without transformation (respectively exports or re-exports) and finally goods in transit between two foreign countries.

General trade includes all goods entering the country as imports, be they stored, distributed for final consumption or transformed as well as, on the export side, all goods of domestic (be they native or transformed imports) or foreign origin (re-exports). Special trade, by contrast, is submitted to customs clearance operations. It includes, on the import side, all goods put at the disposal of importers (destined either for consumption, merchandising or transformation) as well as goods stored in warehouses destined for domestic consumption (with or without transformation); on the export side, exported domestically produced goods as well as exported domestically transformed imported goods. Transit trade should not be included in either category as it does not enter the country (Maizels 1953, 44-47). Nevertheless, in the definition given by the League of Nations, transit trade is explicitly included in general trade (Société des Nations 1928, 10).

Thus, the definitions exposed constitute guidelines rather than universal practice. There are many cases where these definitions have been transgressed, especially before 1914. That is the case in Belgium and the Netherlands, whose customs administration include a large share of their

country's transit trade into its special trade (Horlings 2002, 114). Likewise the distinction between exports and re-exports is not always clear-cut in practice, depending on the permitted extent of "processing." The British customs authorities registered as re-exports (thus included in the general trade figures) and not as exports (part of special trade), tea blends originally imported from abroad and subjected to mixing operations conducted at home (Maizels 1953, 48-49; Giffen 1882, 195-197). Besides, the delimitation between transit and re-exports is not always very straightforward. In the case of 'direct' transit, when cargoes only pass through the country without ever being put at the disposal of importers or stored in warehouses, the goods therein are not even recorded in the customs ledgers; in the case of 'indirect' transit (when goods are stored in bonded warehouses), their exit are regarded as re-exports and included in general trade statistics (Maizels 1953, 49-50). Moreover, some countries do not publish their statistics of special trade. In fact, two standards tended to dominate: a 'Continental' and an 'Anglo-American' standard (Société des Nations 1928). Those countries relying on the former included as imports, "all imports intended for domestic consumption whether direct or taken out of warehouses" and as exports "goods exported from the domestic market, whether purely national in origin or nationalised." The countries relying on the latter included under imports, "direct imports intended for domestic consumption, as well as imports directed to bonded warehouses" and under exports, goods produced on the domestic market as well as those taken out of warehouses, distinguishing between national and foreign goods" (Société des Nations 1928, 8-9). These notions are equivalent more or less to what is understood by general trade. In the first group are France, Belgium, the Netherlands, Germany and Spain; in the second, essentially the United Kingdom and the United States. Extra-European colonies have generally adopted the system of the mother-country. In order to estimate the special trade of the latter group, we subtract the value of export goods classified as foreign total exports and imports (they are regarded as re-exports). This procedure only yields an approximation of special imports (Maizels 1953, 47-48).

The question arises whether to prefer one category to the other. In the League of Nations' *Memorandum* of 1928, the authors assert their preference for special trade figures, because "they are more suitable for international comparisons; besides, they are usually easily accessible at close interval; finally, value figures are always present and broken down in more detail" (Société des Nations 1928, 10). More recently, however, the United Nations have recommended the use of general trade statistics which give a more comprehensive representation of a country's trade compared to special trade numbers (United Nations 1998, 34). The RICardo database has favoured special over general trade when both types of information were available.

Why bilateral flows differ

One of the main richnesses of the RICardo database is that it offers four bilateral flows for every country pair. A trade flow is recorded twice, by the two partner countries in two different sources. It is therefore possible to check on the quality of the data by comparing so-called 'mirror flows'. The term-to-term comparison of mirror flows (exports from A to B as declared by A and imports from B to A as declared by B) can reveal more or less important differences which are the result either of varying procedures for recording flows by country of origin or destination, or methods for estimating trade flows (c.i.f. vs. f.o.b. estimates, trade types), not forgetting the choice of the exchange rate. The difference between mirror trade flows is such an issue that several studies

have been dedicated to its extent and ways to allay its impact. They illustrate perfectly the different cases the scholar might be confronted with (Allen and Ely 1953; Petruzelli 1946).

The recording method is a major source of discrepancy. During the period covered by the database, the geography of bilateral trade flows may have been recorded following one of three conventions then in use. In the first, assessing the direct origin or destination of the goods consists in recording the country from which or into which these are carried directly. The second considers the country of consignment, that is, imports are customarily credited to the country from which last directly consigned and the exports to the country to which first directly consigned. Under the third method, or “production-consumption” method, imports are credited to the country where the good was produced or manufactured in its present form at the time of its entry (country of ultimate origin) and exports to the country of final destination (Durand 1953, 117-126; Don 1968, 89; League of Nations 1928, 25, 28-29). At first sight, the last method appears to be the most sensible one – and the vast majority of trading countries adopted it in the interwar years. However, the second is not without merit, especially for countries carrying an important re-export trade.

France passed successively from the first to the “production-consumption” method in the course of the nineteenth century. Up to 1856, for the trade carried by land, the country of origin or of destination was always the immediate border country, and up to 1869, for the trade carried by sea the country of origin for imports was the last port of call of the carrying vessel and the country of destination for exports that of the first port of call of the carrying vessel. The reform of 1857 accounts, for instance, for the increase of French textile exports to German states that had previously been credited as exports to Belgium.

(Giffen 1882, 189) emphasizes the distortion introduced by the faulty registration of partner countries by customs authorities. He points out that Switzerland was not being registered as such in British trade statistics although trade relations have existed between the two countries for a long time. In our database, which reproduces the information from the British *Annual Statement of Trade*, Switzerland appears as a partner country only from 1906 onwards. But the Swiss Federal Statistical Office started to enter estimates of bilateral trade with Britain as early as 1885.

Durand points to the case of landlocked countries such as Switzerland, Czechoslovakia, Bolivia or Rhodesia, whose external trade was carried in large part over neighbouring countries. Exporters ignorant of or indifferent to the final destination of a particular shipment may have put it on the account of an intermediate country. Trade between the US and Switzerland in the 1930s is cited as an example. Swiss statistics reported a volume of import trade from the US far superior to the export figures to Switzerland quoted in American sources, which attributed these exports destined to this country to Germany, the Netherlands, Belgium, France or Italy. Generally speaking, it is considered that errors about the origin of imports are less likely than errors about the final destination of exports. In fact, to negotiate a bilateral trade agreement, a government needs to know as precisely as possible what part of its imports comes from the concerned partner country; besides, importers can usually ascertain the true primary origin of the goods they import, which is often not the case of exporters who simply do not know the final destination of the goods they ship. Moreover, taxes being mainly levied on imports, the estimation of their quantity and/or value by customs administration is supposed to be more accurate. These are the reasons why import figures are often regarded as more reliable than their export counterparts, and hence preferable (Durand 1953, 123-125).

A case study by (Don 1968, 78-92) of the trade between the United Kingdom and Austria-Hungary between 1895 and 1913 illustrates the problem dramatically. After conversion to a common currency, Austrian statistics appear to record trade flows with the UK five times the size of the flows recorded by British statistics with Dual Monarchy. Possible sources of this bias include: different definitions regarding the type of trade (Austria-Hungary being on a Continental standard does, however, include transit figures into its special trade); differences in the pricing system (since the current prices assigned to exported or imported goods were determined in each country by an official body) and finally by differing methods used for recording countries of origin or destination. The direct origin or destination method was in force in Austria-Hungary until 1891 and until 1904 in the UK. Subsequently Austria-Hungary opted for the production-consumption method while, in 1905-8, the UK adopted the practice of recording the country of consignment, which best suited the type of trade it carried out (the largest re-export trade in the world) (Durand 1953, 121, 126). Only in 1936 did it adopt the production-consumption method.

Many other cases could worsen the picture. But while these criticisms should be kept in mind, they do not undermine the validity or usefulness of a historical database on bilateral trade flows. Despite them, both the IMF and the Barbieri team have adopted the mirror flows strategy with a view not just to discriminate between sources, but in order to supplement missing information.¹² Because of its larger size, the RICardo database provides some ways of mitigating trade data's shortcomings. The approach adopted for the RICardo database has consisted in making no selection, no extrapolation, and no estimation of missing flows. First, the possibility of viewing bilateral trade exchanges over a large time-span facilitates the detection of major inconsistencies. Second, the comprehensive character of a base which offers four bilateral flows for every country pair provides the user with the opportunity to compare data and ultimately select the more reliable figures. The prospective user can then take the option of selecting the associated mirror flow when it is quoted in the base.

RICardo as an exploratory digital tool

Cleaning of entities and currencies

The RICardo database documents a period spanning the beginning of the nineteenth century to the eve of the Second World War and assembles hundreds of thousands of data collected from a variety of sources. In order to be usable, the dataset has been subjected to a standardization process regarding the appellation of the entities and currencies. 'Entities' are the names of the territorial entities recorded under different appellations in the original sources. 'Reporting entities' consist of those entities which collected foreign trade statistics while 'partner entities' are the entities mentioned as the latter's trading partners. The two lists differ substantially because reporting entities are states with customs administration which collect and publish foreign trade figures whereas partner entities may be different kinds of entities: countries, but also harbours, group of countries, or regional areas. A country refers to an entity identified with a COW name (cf. Technical Appendix). This is the major difference with Barbieri's approach which considered only sovereign states as reporting or partner entities. We have developed a standardization process that resulted in the division by 2.8 of the total number of entities referenced in the first version of the database (from 4144 to 1459) (cf. Technical Appendix).

Data from the original sources reproduced in the database are quoted in many different currencies. These quotes are unusable for comparison purposes without an apposite series of exchange rates. We looked for a comprehensive database, which would include all relevant currencies' exchange rate for the period of investigation (1790-1938). Barbieri for her part faced the same problem and she explains how she eventually decided to build up her own exchange rates (Barbieri, Keshk, and Pollins 2008, 15-16). However, her base only covers the post-1870 period and a more limited range of countries than in the present database. In addition, a number of individual flows for a given year are expressed in currencies different from those identified in the RICardo database from the original source. For instance, the Barbieri database provides total trade flows for Argentine, Chile and Columbia in gold pesos. In the sources we consulted, Argentine trade flows were expressed in Pesos Fuertes, Chilean trade in Pesos and Columbian in Piastres. It was impossible to reconstruct the various exchange rates Barbieri applied to the original figures and we decided to assemble our own base of exchange rates for the world's countries from the beginning of the nineteenth century.¹³ The different steps of this operation are described in the Technical Appendix.

The database structure

We have created a relational database (*Figure 1*) that represents all the cleaning and aggregation decisions adopted to provide a usable database. Information on 'entities', 'currencies', and 'sources' appears in specific tables that are linked to the core 'flows' table. Once isolated each type of information is cleaned: the various expressions used in the sources that refer to the same word are reduced to a standardized expression. Extra metadata have been added to the normalized version to complement original information (for example exchange rate to currency). The first version of the database was created and managed under Microsoft Access. When we considered the creation of a web application, we converted it into SQLite format. This conversion process was motivated by the web application technological constraints. It quickly became clear that we needed programmatic data quality checks (scripts which automatically check the database integrity) to support the data cleaning process. This new approach led to a new database format and structure to handle data quality and traceability.

Insert Figure 1 here

As can be seen in *Figure 1*, the database is built around *Flows* data. The surrounding tables used to clean data (entities and currencies) are described in the Technical Appendix.

The 'flows' table

Each trade flow is characterized by the following information:

- *source*: reference of this flow
- *flow*: value of the flow in the original currency
- *unit*: unit of the flow value expressed as a multiplier
- *currency*: original currency name (refers to the *currencies* table)
- *year*: year of the flow
- *reporting*: the entity which reports the flow in its statistics (refers to the *entity_names* table)
- *partner*: the trade partner of the flow (refers to the *entity_names* table)

- *export-import*¹⁴: specifies whether the flow is an import or an export.
- *special_general*: indicates if the flow is of special or general trade or unspecified.
- *species_bullions*: the code is S when trade data include species and bullions; NS when the source indicates that they are not included; null when there is no indication in the source.
- *transport_type*: when indicated in the source, specifies if the merchandise was transported by land, sea, river.
- *statistical_period*: specified when the corresponding year is not a standard civil year.

Two more metadata fields have been coded to identify some complex cases.

The *partner_sum* field is 1 when the flow to and from the partner entity represents a sum of flows and 0 in other cases. Some sources indicate the sum of bilateral flows and such data have been included in the database. This is, for instance, the case of British-Australian trade until 1913. In the dataset we find UK trade with Western Australia, South Australia, Victoria, New South Wales, Queensland and Tasmania in addition to UK trade with Australia (total).

The *world_trade_type* has been added to differentiate three types of total trade corresponding to entity partner RICname 'World'

- *Total_estimated*: data are extracted from sources that only provide countries' total trade. Those data include: re-estimations of a country's total trade over a long period of time (for example Belgium or the Netherlands); data collected from the Mitchell's *International Historical Statistics* volumes; or data coming from the *American Statistical Abstract of Foreign Countries*. National publications of historical statistics are also tagged as 'estimated' as they only supply series of country's total trade.
- *Total_reporting*: in this case, data are extracted from sources that provide countries' total and bilateral trade. The reporting entity type may be: a 'country' with an associated COW code (cf. Technical Appendix); a 'city/part of' a country for which there is no total trade data in the base for the selected year; a 'group' or a 'colonial area' referring to many countries for which there is no total trade in the database for the selected years.
- *Total_subreporting*: data are extracted from sources that provide countries' total and bilateral trade. The reporting entity type is a 'country', a 'city/part of' a country or a 'group' of countries. It is dependent on a larger country for which the database provides total trade data the selected year. This code is used to eliminate duplicates in the calculus of total trade aggregates.

For example: total trade of Danzig (country, COW code = 291) over 1834-1866 (irregular) is coded as *total_subreporting* not to be added to Germany's total trade over these years; in the same manner, Lisbon & Porto's (group) total trade in 1840 is coded as *total_subreporting* not to be added to Portugal's total trade this year; total trade of Bahia (city/part of) over 1831-1874 is coded as *total_subreporting* not to be added to Brazil's total trade-

The 'Sources' and 'source_types' tables

The *sources* table lists all the sources contained in the database. A detailed reference to the original document allows the user to check the work of transcription. It is also a way of inviting the user to suggest better alternative sources. The *source_types* table then classifies the whole set of sources according to three categories: 1. 'primary' sources are compilations of customs returns published by national agencies; 2. 'secondary' sources include: either international compilations of

trade primary sources which were in bygone days devised by individuals and government agencies and are today a function of international organisations, or national compilations of national statistics that include foreign trade statistics (mainly statistical yearbooks); 3. ‘estimation’ sources are works that only provide country’s total trade. They include re-estimations of long-term series of a country’s total trade, compilations of historical series of world countries’ total trade, or national publications of historical statistics.

As described in this section, we use a relational database to support and document the data cleaning process. Its structure has been designed to allow the best transparency about the choices made. In order to simplify future data updates (correcting errors or adding new data), the relational database can be exported as a series of Comma Separated Value (CSV) tables under a version-controlled tool (git) which allows us to keep track of every change.

RIC visualizations

To foster the use of the RICardo database, we have developed a web application where the user can explore the dataset through data visualizations (Girard et al. 2016). It is a pioneering work that aims at experimenting new ways of exploiting historical trade statistics and opening access to such data to a broader public, beyond the boundaries of academic circles. We hope that this project will contribute to illustrating the invaluable interest of working with historical data series. The database, available in open-access¹⁵, will be publicly released in 2017 upon the start of an international conference marking the two-hundredth anniversary of the publication of David Ricardo’s *Principles*. Here is a presentation of the exploration interface. It offers four levels of entry from global to local.

Metadata view

The metadata view provides an overview of all the information available in the database. A first visualization reports the number of flows by year. Annual variation is a reflection of either an increase in the number of reporting entities or in the number of partner countries by reporting entity, that can be due to the sources used. The visualization below is built around a detailed reporting/year matrix where the user can find six variables for each reporting/year:

- World partner: as there are different ways of calculating world trade series, it specifies what type of ‘world’ (estimated, reported, sum) is used to calculate the *world_best guess* (cf. ‘world view’ below);
- Number of partners: information on the total number of partners by reporting/year is synthesized by graduated color. The distribution of the number of partners by continent can be seen by hovering the mouse pointer over a matrix cell;
- Bilateral rate: it informs on the proportion of partners for which mirror flows can be estimated (cf. ‘bilateral view’ below);
- Source type: it provides an overview of primary/secondary sources by reporting. The type ‘estimation’ means that only total trade data are available for this reporting/year;
- Reporting type: it can be a country, a city/part of, a colonial area, or a group. This selection shows that the last three types correspond to scattered data in the earlier period;

- Reporting continent: Africa, America, Asia, Europe, Oceania, World. When the ‘Reporting type’ is a group of countries from different continents, the associated continent is ‘world’.

World view

The *world* view focuses on total trade data. Its purpose is to provide an estimation of world exports and imports over the long run and to compare the relative share of countries in world trade over a selected period.

When creating this view, we were confronted with two sorts of problems. The first one is the variation in the number of reporting entities which constitute the world. It must be emphasized that the “*world trade*” curves depicted in this view aggregate a variable number of reporting entities from year to year. This number is indicated in a bar chart. The user also has the opportunity of estimating homogeneous world trade series by downloading an appropriate dataset. The second problem relates to the different ways of estimating total trade. It has been pointed out that different types of total trade can be differentiated in the RICardo database. *Total_estimated* emanates from sources that only provide total trade data. *Total_as reported* coincides with the ‘Total’ line appearing in trade statistics. *Total_sum partner* is total trade as obtained from the addition of all partner entities available in the database. *Total_best_guess* selects the “best” available data for a reporting/year on a priority basis: *Total_estimated* is first chosen; if not available, *Total_as reported*; and the latter option is *Total_sum partner*. As mentioned above, it is planned to enrich this comparison with the inclusion of *Total_FedericoTena* series, which are the most elaborated estimations of total trade to this day.

Country view

This view concentrates on the analysis of a selected reporting country. The objective is to facilitate comparisons between its partners by ordering them according to an average share (annual average calculated for available years over the selected period), representing annual trade balances, and depicting the growth in trade shares for selected partners.

This view is very useful to spot unexpected problems with entity names. In fact, when a partner is identified as a city, a part of an entity, or a group of entities in trade statistics, we have not tried to reallocate the flow to one or more ‘*country*’ type partners. It may thus happen that information on bilateral trade of two countries appear incomplete in the visualization when the name of the partner entity is changing. Let us take the example of Anglo-American trade. British trade statistics report trade with the United States of America from 1831 onwards, except over 1871-1905 where the American partner is recorded under two different partner names: the United States_Atlantic coast and the United States_Pacific coast.

Insert Figure 2 here

Furthermore, the country view can be used to visualize the impact of a change in the method of recording trade statistics. The example of France, which introduced in 1857 a reform in the way it records countries of origin and destination for trade carried by land, was given above. The re-

searcher can carry out a first assessment of the impact that this reform entailed in trade data by looking at the variation in the trade balance with border partners in 1858. Sudden change appears for Austria, Belgium, Germany/Zollverein, Mecklenburg-Schwerin, and Spain that requires a more thorough investigation.

Insert Figure 3 here

Bilateral view

The purpose of this view is to interrogate a pair of *reporting* countries by using a representation of their mirror flows. As outlined above, one of the strengths of the RICardo database is the possible existence of two data for one bilateral flow. However, for a number of reasons given in a preceding section, the two recordings are rarely equal. Thanks to the calculation of an indicator (Dedinger 2012, 1050-1053), the bilateral view provides a visualization of the discrepancy between mirror flows and of its fluctuations over the selected period.

A significant deviation between Austrian and British trade statistics between 1895 and 1913 has been pointed out by (Don 1968). This can be easily verified by looking at the Anglo-Austrian country pair from the two bilateral points of view. With the United Kingdom as reporting – i.e. Anglo-Austrian trade as charted by British statistics - a great discrepancy in the mirror flows is actually observed due to a relatively important under-estimation of the value of trade in British statistics. With Austria-Hungary as reporting, the problem persists but it seems reduced when the discrepancy is computed in respect of Austrian statistics. It is recalled that the sharp decline in the mirror flows discrepancy in 1906 is due to the adoption of the country of consignment method by the United Kingdom in 1905. The bilateral view is thus a useful tool to facilitate the detection of problems in trade statistics and to identify the reporting countries that perform better in the quality of their trade statistics.

Insert Figures 4 and 5 here

Eventually, the user can download the set of data produced to support the visualizations as CSV (comma separated value) files.

Conclusion

Describing the various phases of our work from the first conception to its final outcome, it is intended to provide the reader/user with a better knowledge of historical trade statistics and a guide to use the RICardo database. The creation of this innovative tool was made possible thanks to a close collaboration between researchers with complementary expertise and we would like to

perpetuate this spirit of cooperation by openly sharing our work and inviting the user to contribute to the development of the project.

It is hoped that the RICardo project can serve a variety of purposes, scientific as well as educational. In the scientific arena, RICardo can help to renew research on trade history in at least three different ways. The analysis of trade globalization – causes or effects – can be revisited thanks to extended trade data series. In a first experiment using an earlier version of the RICardo database, (Accominotti and Flandreau 2008) put into question the economic impact of the wave of bilateral trade treaties – the so-called “Cobden-Chevalier network” – that spread over Europe after the signature of the Cobden-Chevalier Treaty in 1860 and challenge the conventional view that it was at the root of the nineteenth-century trade liberalization. Furthermore, the database may encourage research on trade globalization in fields that tend to be abandoned by economic historians, particularly quantitative studies of a country’s trade history. Finally, the RICardo database can be a good instrument to help move the secular debate on the reliability of trade statistics.

The online publication of the base is a significant achievement but it does not put an end to the project. Apart from the periodical updating of data, further improvements in the visualizations are planned in the future. One is to take into account the political dimension of trade by focusing on bilateral trade with/of colonies and federal states. It implies that we add time information in the codification of the entities to track the development of their status over time. Another improvement is to offer a new visualization on the trade of continents distinguishing intra- and extra-trade. Eventually, a cartographic representation should complement the visualization of historical trade data.

REFERENCES

- Accominotti, O., and M. Flandreau 2008. Bilateral Treaties and the Most-Favored-Nation Clause: the Myth of Trade Liberalization in the Nineteenth Century. *World Politics* 60:147-88.
- Alex, J. 1861. Memorandum indicating some of the causes to which the discordancy of the commercial statistics of various countries is chiefly to be attributed. *Journal of the Statistical Society of London* 24:230-31.
- Allen, R.G.D., and J. E. Ely, eds. 1953. *International Trade Statistics*. New York: John Wiley & Sons.
- Bairoch, P. 1973. European Foreign Trade in the XIXth Century : The Development of the Value and Volume of Exports (Preliminary Results). *The Journal of European Economic History* 2:5-36.
- Bairoch, P. 1974. Geographical Structure and Trade Balance of European Foreign Trade from 1800 to 1970. *The Journal of European Economic History* 3:557-608.
- Bairoch, P. 1976. *Commerce extérieur et développement économique de l’Europe au XIX^e siècle*. Paris: Ecole des Hautes Etudes en Sciences Sociales.
- Barbieri, K., 1996. *Economic Interdependence and Militarized Interstate Conflict, 1870-1985*. Ph.D. Dissertation Binghamton University.
- Barbieri, K. 2002. *The Liberal Illusion: Does Trade Promote Peace?* Ann Arbor: University of Michigan Press.
- Barbieri, K., O. Keshk, and B. M. Pollins. 2008. *Correlates of War Project Trade Data Set Codebook*, Version 2.01.

- Barbieri, K., O. Keshk, and B. Pollins. 2009. TRADING DATA: Evaluating our Assumptions and Coding Rules. *Conflict Management and Peace Science* vol. 26:471-91.
- Barbieri, K., and O. Keshk. 2012. *Correlates of War Project Trade Data Set Codebook, Version 3.0*. <http://correlatesofwar.org>
- Clark, G. N. 1938. *Guide to English commercial statistics 1696-1782*. London: Offices of the Royal Historical Society.
- Charles, L., and G. Daudin. 2011. Le Bureau de la balance du commerce. *Revue d'Histoire Moderne et Contemporaine* 58:128-155.
- Charles, L., and G. Daudin (Ed.). 2015. Eighteenth-Century International Trade Statistics. Sources and Methods. *Revue de l'OFCE* 140.
- Denzel, M. 2010. *Handbook of World Exchanges Rates, 1590-1914*. Surrey: Ashgate Publishing, Ltd.
- Don, Y. 1968. Comparability of International Trade Statistics: Great Britain and Austria-Hungary before World War I. *Economic History Review* 21:78-92.
- Durand, E. D. 1953. Country Classification. In *International Trade Statistics*, edited by R. G. D. Allen, and J. E. Ely, 117-29. New York: John Wiley & Sons.
- Escaith, H. 2015. Past and present issues in trade statistics. *Revue de l'OFCE* 140: 39-51.
- Federico, G., and A. Tena-Junguito. 1991. On the Accuracy of Foreign Trade Statistics (1909-1935): Morgenstern Revisited. *Explorations in Economic History* 28:259-273.
- Federico, G., and A. Tena-Junguito. 2016. World trade, 1800-1938: a new data-set. *EHES Working Papers in Economic History* 93.
- Girard, P., B. Dedinger, D. Ricci, B. Ooghe-Tabanou, M. Jacomy, G. Plique, and G. Tible. 2016. RICardo Project: Exploring XIX Century International Trade. Paper presented at Digital Humanities 2016, Krakow, Poland, July 13–15, 2016.
- Gowa, G., and R. Hicks. 2013. Politics, Institutions, and Trade: Lessons of the Interwar Era. *International Organization* 67: 439-467.
- Gowa, G., and R. Hicks. 2015. Commerce and Conflict: New Data about the Great War. *British Journal of Political Science*. DOI: <http://dx.doi.org/10.1017/S0007123415000289>, 22 pages.
- Horlings, E. 2002. The International Trade of a Small and Open Economy. Revised Estimates of the Imports and Exports of Belgium, 1835-1990. *NEHA-Jaarboek* 65:110-42.
- Hugot, J. 2015. *Trade Costs and the Two Globalizations*. http://www.econ.itu.edu.tr/documents/activity20142015/Paper_Hugot.pdf
- International Monetary Fund. 1993. *A Guide to Direction of Trade Statistics*. Washington, D.C.: International Monetary Fund.
- Jacks, D. S., C. N. Meissner, and D. Novy. 2011. Trade booms, trade busts, and trade costs. *Journal of International Economics* 83: 185-201.
- League of Nations. 1927. *International Statistical Yearbook 1926*. Geneva: League of Nations.
- League of Nations. 1945. *Industrialisation and Foreign Trade*. Geneva: League of Nations.
- Lewis, W. A. 1952. World Production, Prices and Trade, 1870-1960. *The Manchester School of Economic and Social Studies* May:105-138.
- Lewis, W. A. 1981. The Rate of Growth of World Trade, 1830-1973. In *The World Economic Order. Past and Prospects*, edited by S. Grassman, and E. Lundberg, 11-81. London: Macmillan.
- Maddison, A. 1962. Growth and Fluctuation in the World Economy, 1870-1960. *Banca Nazionale del Lavoro Quarterly Review* 61:127-195.

- Mc Culloch, J. R. 1841-1866. *A Dictionary, geographical, statistical, and historical of the various countries, places, and principal natural objects in the world*. London: Longman, Ome, Brown, Green and Longmans.
- Maizels, A. 1953. Coverage. In *International Trade Statistics*, edited by R. G. D. Allen, and J. E. Ely, 44-50. New York: John Wiley & Sons.
- Marín, A. C. 2005. *Geographical effects on the accuracy of textile trade data: an international approach for 1913*. Barcelona: UPF Economics and Business Working Paper n° 821.
- Mitchell, B. R. 1975. *European Historical Statistics, 1750-1970*. London: Macmillan Press.
- Mitchell, B. R. 2007. *International Historical Statistics, Europe: 1750-2005*. New York: Basingstoke.
- Mitchell, B. R. 2007. *International Historical Statistics, the Americas: 1750-2005*. New York: Basingstoke.
- Mitchell, B. R. 2007. *International Historical Statistics, Africa, Asia & Oceania: 1750-2005*. New York: Basingstoke.
- Mitchener, K. J., and M. Weidenmier. 2008. Trade and Empire. *The Economic Journal* 118: 1805-1834.
- Morgenstern, O. 1963. *On the Accuracy of Economic Observations*. Princeton: Princeton University Press.
- Morgenstern, O. 1972. *L'illusion statistique. Précision et incertitude des données économiques*, Paris: Dunod.
- Mulhall, M. G. 1884-1899. *A Dictionary of Statistics*. London: Routledge and sons.
- Pascali, L. 2014. *The Wind of Change: Maritime Technology, Trade and Economic Development*. <http://hdl.handle.net/10230/22608>
- Petruzelli, N.M. 1946. *Some Technical Aspects of Foreign Trade Statistics with Special Reference to Valuation*. Washington, D.C.: The Catholic University of America Press.
- Platt, D.C.M. 1971. Problems in the Interpretation of Foreign Trade Statistics before 1914. *Journal of Latin American Studies* 3:119-30.
- Schneider, J., O. Schwarzer, and M. A. Denzel. 1991. *Währungen der Welt I. Europäische und Nordamerikanische Devisenkurse, 1777-1914. Teilband I*. Stuttgart: F. Steiner.
- Schneider, J., O. Schwarzer, and M. A. Denzel. 1991. *Währungen der Welt I. Europäische und Nordamerikanische Devisenkurse, 1777-1914. Teilband II*. Stuttgart: F. Steiner.
- Schneider, J., O. Schwarzer, and M. A. Denzel. 1991. *Währungen der Welt I. Europäische und Nordamerikanische Devisenkurse, 1777-1914. Teilband III*. Stuttgart: F. Steiner.
- Schneider, J., O. Schwarzer, and M. A. Denzel. 1992. *Währungen der Welt IV. Asiatische und australische Devisenkurse im 19. Jahrhundert*. Stuttgart: F. Steiner.
- Schneider, J., O. Schwarzer, and M. A. Denzel. 1997. *Währungen der Welt VII. Lateiamerikanische Devisenkurse im 19. Und 20. Jahrhundert*. Stuttgart: F. Steiner.
- Schneider, J., O. Schwarzer, and M. A. Denzel. 1994. *Währungen der Welt VIII. Afrikanische und levantinische Devisenkurse im 19. und 20. Jahrhundert*. Stuttgart: F. Steiner.
- Société des Nations. 1928. *Mémoire sur le commerce international et sur les balances des paiements 1912-1926, vol II*. Genève : Société des Nations.
- Svennilson, I. 1954. *Growth and Stagnation in the European Economy*. Geneva: United Nations.
- United Nations. 1998. *International Merchandise Trade Statistics: Concepts and Definitions*. New York: United Nations.

Technical Appendix

A – The “entities’ tables

The non-normalized dataset contained thousands of different entity names, which had to be standardized in order to obtain coherent series of data over the whole period. The most frequent cases of variation were: the same entity appears for the same year under different names in different publications; the same entity changes names during the period under consideration; the trading entity does not correspond to an internationally recognized state but describes a locality or a limited section thereof; the trading entity is not defined precisely. Here are the main steps followed in the standardization operation:

Cleaning the entity names

The first task has consisted in reformatting the names of reporting and partner entities, the outcome of which appears in the *Entity_names* table. A number of spelling and other mistakes have slipped through in the original sources, and entity names have been entered in different languages and under varying appellations. The *Entity_names* table converts the ‘*Original Name*’: 1. into a corrected French ‘*Name*’; 2. into a standardized ‘*RICname*’, as can be seen in *Table 1*.

Insert Table 1 here

RICname entities

The *RICname* is in English. It is partly established after the COW project country list as we explain below. Each *RICname* is identified by a number of three variables: type, continent, COW code:

- the ‘*type*’ variable indicates if the entity is: a city/part of a country, a colonial area, a country, a geographical area, or a group of countries.
- the ‘*continent*’ column indicates where the (reporting or partner) entity is located. When the entity is a ‘group’ or a ‘geographical area’ corresponding to a multiple of entities located on different continents, it is identified by the continent ‘World’. The five major continents (Africa, America, Asia, Europe, Oceania) are present in the database plus “continents” pertaining to sea areas (Adriatic, Antarctic, Atlantic Ocean, Baltic, Mediterranean, Pacific). Turkey, when non-specified in Europe or in Asia, has been located in Asia; Russia/USSR, when non-specified in Europe or in Asia, has been located in Europe. The user can change these options if necessary.

The ‘country’ type *RICname* is defined by referring to the *Correlates of War (COW)* project. Each RICardo entity that is referenced in the COW list is identified as a ‘country’ type and is given a COW code. The COW project, initiated in 1963 by American political scientists, has collected quantitative information about armed conflicts in the post-Napoleonic period and resulted into the constitution of several databases, two of which are concerned with the definition and inventory of state entities.¹⁶ Thus, the *State System Membership List* contains the list of all entities which have enjoyed the internationally recognized status of sovereign state as of 1815.¹⁷ The *Colonial/Dependency Contiguity Data* variable identifies every contiguity situation (land or river

boundaries, or bodies of water) of political entities of the international system (sovereign states, colonies and dependencies) and leads to drawing up a subsidiary list of colonies and dependencies belonging to sovereign states. A ciphered code is attributed to each of these entities with dates of changes of political status. All this information (including entity name and code, political status, and relevant time periods) is a 50 odd-page document entitled *Entities.pdf*; that served as a basis to define the name and code of each ‘country’ type *RICname*.

For those entities, which do not show up in the COW database – i.e. ‘city/part of’, ‘colonial area’, and ‘geographical area’ – new English entity names have been created without a numerical code. Furthermore, new entity names and new codes have been created for three ‘countries’ that are not included in the COW database:

Kingdom of Sardinia (325S): created in 1720, this political entity (assimilated to Italy/325 in the COW list) consisted, before the unification of Italian states in 1861, of Savoy, Piedmont (Turin), Aosta, Nice and the island of Sardinia.

Prussia (255P) and Germany (Zollverein) (255Z): Prussia is assimilated to Germany/255 in the COW list. But before the foundation of the German empire (1871), several German entities appear in the RICardo list of partner names – Prussia, German states, Germany and German Zollverein – which do not correspond to similar territories. Thus, new country names have been created for Prussia and the German Zollverein (which may be considered as an economic union). We have translated ‘German states’ and ‘Germany’ into ‘Germany’ although ‘Germany’ did not actually exist before 1871.

Insert Table 2 here

B – Currency conversion

The conversion of the values expressed in many different currencies was carried out in two successive steps: 1. Drawing a conversion table exhibiting standardized currency names; 2. Drawing a corresponding table for each currency’s exchange rate to the British pound in any given year. Our *Exchange Rate* table includes all the relevant exchange rates necessary for converting the original quotations into pound sterling. This is an arbitrary choice, but the pound sterling estimates can in turn be easily converted into US dollars or French francs with the appropriate converters.

The ‘currencies’ table

Before dealing with arithmetic, it was necessary to standardize the currency names found in the original database in order to link each currency to a single entity for a selected period. Instances can be found in the original of different appellations for the same currency, in the same country for an identical year as scripture, language or type of document varied from one sources to the next. After due verification, the *currencies* table links every original currency name to its standardized version for a given year. Here is an example:

Insert Table 3 here

The ‘exchange rates’ table

In the French *Annales du commerce extérieur* and the British *Statistical Abstract for the Foreign Countries*, flows are quoted either in national currency and/or in francs or pounds. When the original figures expressed in national currency were converted, the exchange rate used was

always based on the gold parity. By contrast, when the information was available, we have used current exchange rates. When it was not, the gold parity was used. For this purpose the main sources tapped were: *Annales du Commerce extérieur*, *Statistical Abstract for the Foreign Countries*, the *Währungen der Welt* collection,¹⁸ the *Montevideo-Oxford Latin American Economic History Data Base* (<http://www.lac.ox.ac.uk/moxlad-database>), *Statesman's Yearbook*. For the US dollar-pound exchange rate, the current exchange rate given in (Denzel 2010) has been preferred except during the years 1862-64 for which the dollar-sterling parity is used.

From the list of standardized currency names, corresponding exchange rates in pound sterling (NCU per £) have been added on an annual basis with the mention of the original source in the *Exchange Rate* table. This table includes ca. 15000 exchange rate quotations. Each row refers through the *source* field to the *sources* table which lists the complete reference to sources used. Despite this relative wealth of information, this particular database is liable to be supplemented when need be.

This *Exchange Rate* table will be made available for download with the rest of the database in 2017.

Figure 1. RICardo database schema

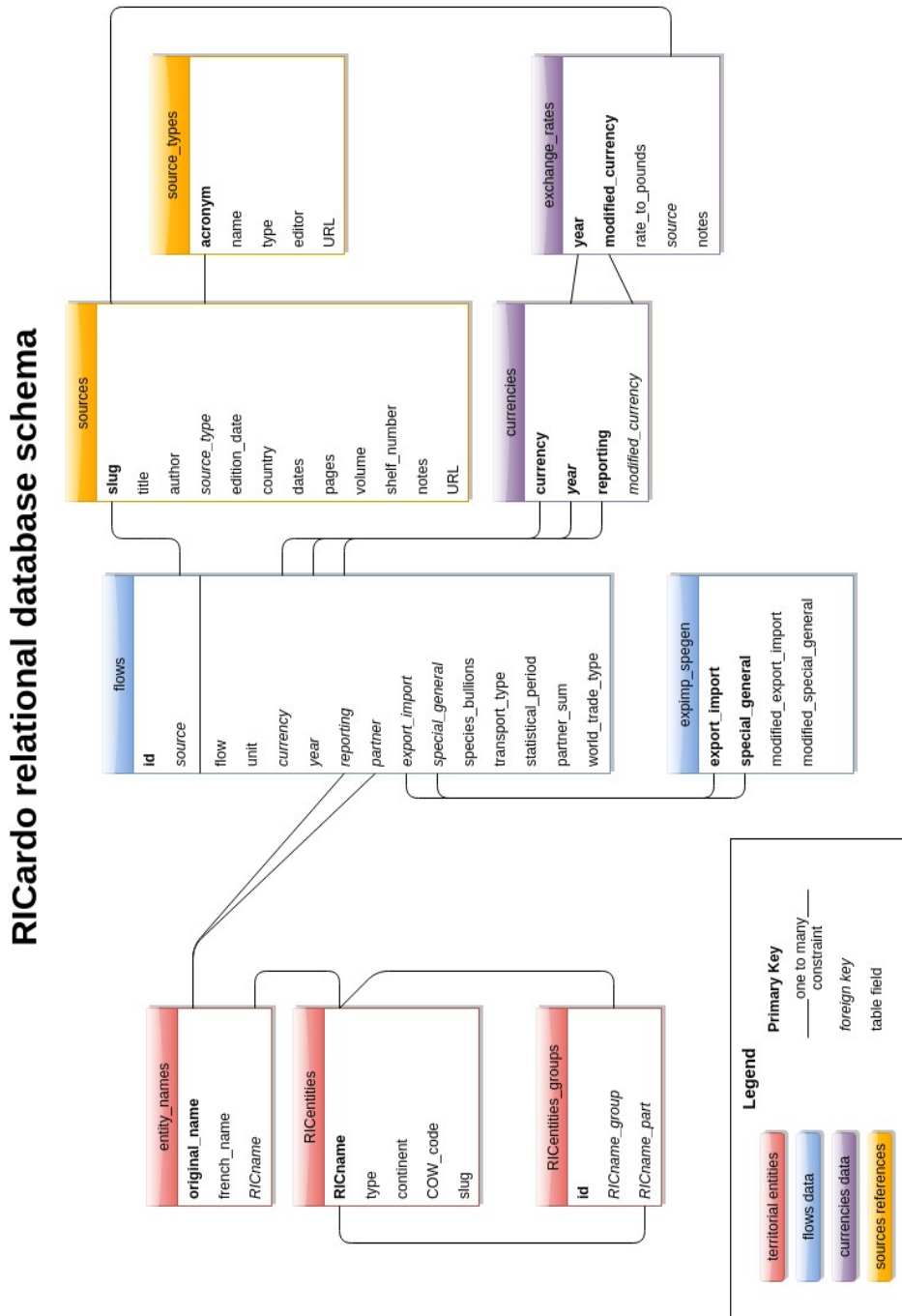


Figure 2.



Figure 3.

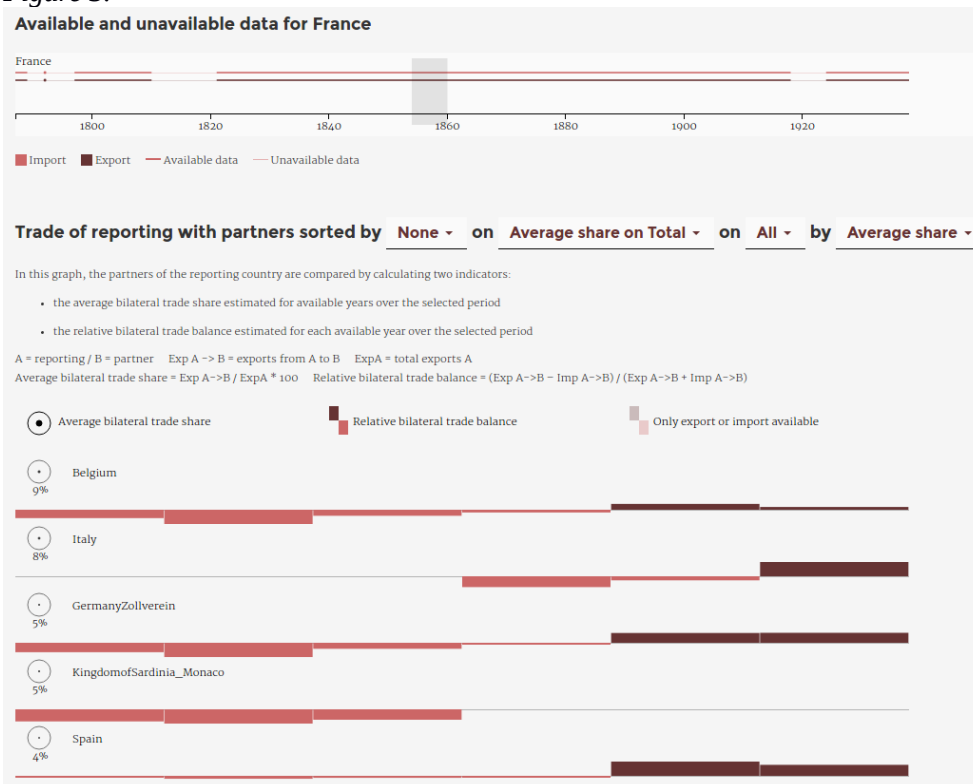


Figure 4.

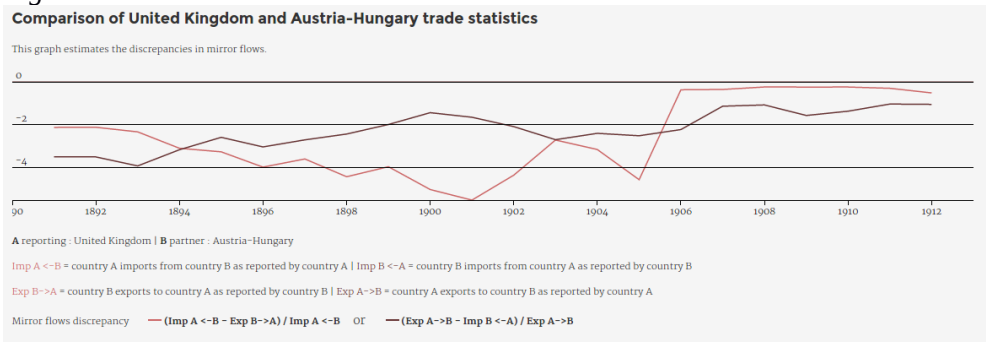


Figure 5.



Table 1. Normalization of entity names

Original name	French name	RICname
Afrique, Cap Bonne Espérance	Colonie du Cap	Cape Colony
British Possessions in South Africa : Cape of Good Hope	Colonie du Cap	Cape Colony
Cabo de Buena Esperanza	Colonie du Cap	Cape Colony
Cap	Colonie du Cap	Cape Colony
Cap (Colonie du)	Colonie du Cap	Cape Colony
Cap Bonne Espérance	Colonie du Cap	Cape Colony
Cap de Bonne Espérance	Colonie du Cap	Cape Colony
Cap de Bonne-Espérance	Colonie du Cap	Cape Colony
Cape Colony	Colonie du Cap	Cape Colony
Cape Colony (Colonia del Cabo)	Colonie du Cap	Cape Colony
Cape of Good Hope	Colonie du Cap	Cape Colony
Cape of Good Hope (South Africa)	Colonie du Cap	Cape Colony
Cape of Good Hope including Kaffraria (South Africa)	Colonie du Cap	Cape Colony
Colonie Cap	Colonie du Cap	Cape Colony
Colonie du Cap	Colonie du Cap	Cape Colony
South Africa - Cape of Good Hope	Colonie du Cap	Cape Colony
South Africa : Capte of Good Hope	Colonie du Cap	Cape Colony
South Africa, Cape of Good Hope	Colonie du Cap	Cape Colony
Union of South Africa : Cape of Good Hope	Colonie du Cap	Cape Colony
Union of South Africa : Cape of Good Hope§	Colonie du Cap	Cape Colony

Table 2. Codification of RICentities

RICname	Type	Continent	COW code
Cape Colony	Country	Africa	561
Dalny	city/part_of	Europe	
Dalny & Korea (Chosen) & Vladivostok	Group	World	
Damietta	city/part_of	Africa	
Danish America	colonial_area	America	
Danish Colonies	colonial_area	World	
Danish Colonies & Denmark	Group	World	
Danish Europe	colonial_area	Europe	
Danzig	Country	Europe	291
Danzig & Memel	Group	Europe	
Danzig & Poland	Group	Europe	
Davis Strait	geographical_area	America	
Davis Strait & Greenland	Group	America	
Denmark	Country	Europe	390
Zimbabwe (Rhodesia) (Southern Rhodesia)	Country	Afrique	552

Table 3. Standardization of currency names

Original Currency	Year	Reporting	Modified currency
Bolivares	1830	Venezuela	Venezuelan Bolivar
Bolivian Boliviano	1935	Bolivia	Bolivian Boliviano
Colones	1924	Costa Rica	Costa Rican Colon
Colones	1931	Salvador	Salvadoran Colon
Couronnes	1874	Suède	Swedish Krone
Couronne (Kroner)	1880	Norvège	Norwegian Krone
Couronnes	1875	Danemark	Danish Krone
Drachmes	1845	Greece	Greek Drachma
Drachmas	1890	Grèce	Greek Drachma
Kroner	1924	Danemark	Danish Krone
Kroner	1890	SWE	Swedish Krone
Kroner	1872	Norway	Norwegian Krone
Pesos	1862	Peru	Peruvian Peso
Pesos	1925	Cuba	Cuban Peso
pesos	1926	Uruguay	Uruguayan Peso
pesos	1901	Salvador	Salvadoran Peso

NOTES

1 <http://ricardo.medialab.sciences-po.fr/> (for reviewers only).

2 “Ransacking these sources is very time-consuming, and very frustrating; and it is hardly worthwhile to spend a day in the library discovering that a particular country’s exports were valued at \$5 million in 1855 (0.2 per cent of the world total).”

3 In his later OECD historical publications, Maddison recycled his earlier exports estimates of 1962.

4 Austria, Belgium, Canada, Denmark, France, Germany, Ireland, Italy, the Netherlands, Norway, Sweden, Switzerland, United Kingdom, United States.

5 In the two quoted articles, no systematic reference to the sources used is made; the author explains the adjustments he made to the data. For an exposition of the data he used on the nineteenth century, see (Bairoch 1976).

6 All the material gathered by Bairoch is stored at the *Institut d’histoire économique et sociale* in Lausanne.

7 Barbieri uses the classification defined by the Correlates of War project, which we have also followed. Cf. below.

8 A polity is defined as “a political entity which can register its own trade and is registered as a separate entry in the trade statistics of other polities, including colonies.”

9 This is a more than hundred-year-old debate philosophical as well as technical that can be summarized with the most representative quotations on both sides: “*It would seem, finally, that comparison of trade statistics, for a historical analysis of economic relations between two countries, must be abandoned. The selection of the statistical data of one of the two countries under consideration must depend upon the particular purposes of the research, and must take into account the methods of compiling foreign trade statistics used by each of the two countries*” (Durand 1953, 92).

“*Let us consider what impression will probably remain on the mind of anyone who has carefully examined [commercial statistics], and who has no extraneous evidence to guide him in*

estimating the degree of authority to be assigned to any of the statements, which they embody. Will he not conclude that he must refuse his assent to all these conflicting statements alike, and submit, as he best can, to remain in ignorance of the facts which each of them professes to set before his eyes? [...] We trust we shall be able to show that the disagreement between these statements, embarrassing as its existence doubtless is, is not inconsistent with the essential veracity of each, being for the greater part produced by a difference in the form and arrangements of the records from which they are abstracted – an apparent not a real contradiction” (Alex 1861).

10 Other titles include: *Statistical Abstract for the several British self-governing Dominions, Colonies, Possessions, and Protectorates in each year ...; Statistical Abstract for the several British Overseas Dominions and Protectorates; Statistical Abstract for the British Empire for each of the years...*

11 Other titles include: *Mémorandum sur le commerce international et sur les balances des paiements, Statistiques du commerce international.*

12 Missing data cells may be filled with data reported by trade partners adjusted by a uniformly applied percentage of 10 % to allow for the cost of freight and insurance (IMF 1993, 9-10). Given the gaps in the trade statistics of many countries, especially before 1900, Barbieri supplements the missing data with information retrieved from the statistics of more developed countries assuming the equivalence of mirror flows (Barbieri and al. 2008, 14).

13 Global Financial Database (<https://www.globalfinancialdata.com/>) provides daily data on currency fluctuations going back to the 1920s and yearly exchange rates through the nineteenth century. But apart from some problems on the reliability of data, this database could not meet all of our exchange rate needs. Furthermore, it is not freely accessible whereas we want to give the user free access to all our data.

14 Both `import_export` and `special_general` original value in source are corrected to a standard denomination by the dedicated table `expimp_spegen`.

15 Cf. note 1.

16 More information available on www.correlatesofwar.org. The relevant database is: [Correlates of War Project](#). 2011. "State System Membership List, v2011." Correlates of War 2 Project, Colonial/Dependency Contiguity Data, 1816-2002, Version 3.0.

17 Defined as sovereign states are: 1. Before 1920 all political entities with a population of at least 500,000 people having entertained diplomatic relations (in the person of at least a chargé d'affaires) with Britain and France, and 2. after 1920, all country members of the League of Nations (later, of the United Nations) or alternatively all entities with a population of at least 500,000 and diplomatic representation with at least two "great powers" (including according to COW: Germany, China, the US, France, Italy, Japan, the UK and Russia-USSR).

18 A series of 11 volumes edited by Jürgen Schneider, Oskar Schwarzer & Markus A. Denzel (Stuttgart, Steiner, 1991-9) covering world exchange rates going back to the seventeenth century the world over. Supplemented by a new edition by M. Denzel (2010).