

RICARDO WORKSHOP 2017

The first RICardo workshop was organized in Paris on 12 May 2017 to celebrate the official launching of the RICardo website on the bicentenary of David Ricardo's *Principles of Political Economy and Taxation*. It gathered the members of the newly created RICardo committee* to discuss a number of methodological questions pertaining to historical trade statistics and their exploitation with digital technology. This meeting was an opportunity to exchange views on challenging questions that each trade database builder has to face and answer. A list of discussion points** was submitted to the committee. The outcome of our discussions is intended to help both the inexperienced and the scholar user to better know and utilize trade data and the RICardo website.

*RICARDO COMMITTEE

Katherine Barbieri (University of South Carolina / COW Trade Data Set and *Barbieri* database)

Bertrand Blancheton (Université Bordeaux / *Montesquieu* database)

Guillaume Daudin (Université Paris-Dauphine / *Toflit18* and *RICardo* projects)

Béatrice Dedinger (SciencesPo, Centre d'histoire de Sciences Po / *RICardo* Project)

Hubert Escaith (WTO, Former Chief Statistician)

Giovanni Federico (Università di Pisa / *Federico-Tena* database)

Paul Girard (SciencesPo, médialab / *RICardo* Project)

Kevin O'Rourke (All Souls College Oxford / English_British_UK trade statistics 1700-1899. Not present at the workshop)

Donato Ricci (SciencesPo, médialab / *RICardo* Project)

Antonio Tena (Universidad Carlos III de Madrid / *Federico-Tena* database)

**LIST OF DISCUSSION POINTS

Why are we interested in historical trade databases?

No "universal" trade database

Importance of the quality of the data

Sharing data with the public

Visualizing the data

Total trade versus bilateral trade

Different estimations of total trade

Why do multiple estimations exist for total trade?

The adjustment procedure in 'World Federico-Tena' series and bilateral trade series

Currency conversion

Exchange rate databases

Problems with historical exchange rates

Missing data

What are the problems with missing data?

Are bad data better than no data?

Trade valuation

How are trade values derived?

A case study to better understand trade valuation: Franco-Belgian trade in 1910

Mirror flows

The RICardo bilateral view
Potential use of RICardo's mirror flows

Entities

The RICentity table
RICentities' flaws
Is the creation of a 'perfect' database a practicable option?

Future prospects

New features on the website
Updating the database
User interface

WHY ARE WE INTERESTED IN HISTORICAL TRADE DATABASES?

Going back to fundamentals, the committee first questioned the motivations of each of its members in constructing and exploiting trade databases. It is interesting to point out that very similar objectives were expressed by each participant, giving a spirit of cohesion to an eclectic group of people with diverse skills and origins. Why are we invested in historical trade databases? There is one precision and three major objectives.

- No "universal" trade database

RICardo is a trade database containing total and bilateral trade statistics. Its spatial and temporal coverage are not limited *a priori*, but it does not seek to extend further to trade by products. It complements other historical trade databases (Barbieri, Fouquin-Hugot, Gowa-Hicks) as well as projects currently underway to reconstitute historical trade databases at a national level (such as TOFLIT18, Duanaire, Sound Toll Registers, digitalarkivet).

- Importance of the quality of the data

We all strive to improve data quality. Like all data, trade statistics, particularly historical trade statistics, are problematic and scholars are divided about the best way to handle different problems. We try to improve the quality of trade data by giving careful attention to the choice, cleaning and checking of data (trade data as well as exchange rate data). Secondly, we seek to alert the users to the potential limitations of trade data. Researchers that use trade data in their empirical research are not always aware of the problem of data quality. We would like to reinforce the information for the user, what can be easier through the channel of a website.

- Sharing data with the public

A common objective is to give open access to trade statistics data to the widest possible audience. Our primary target audience is researchers, but the website should be useful to students and teachers from high school to university. New ways of exploiting these statistics could thus be explored, away from traditional gravity models. We also think that trade statistics can be used in fun and exciting ways.

- Visualizing the data

Adapting to new technologies, researchers now have the opportunity to add a new dimension to trade databases by cooperating with developers and designers. Data visualization opens new opportunities for research and teaching by providing instant and detailed access to the

data. It is particularly useful to show historical trends in international trade relationships and to help us understand them. Our common goal is innovation, creativity, and service to the user.

TOTAL TRADE VERSUS BILATERAL TRADE

The total trade of a country can be estimated in different ways. Several estimations are provided on the RICardo website to make the user aware of these differences and more curious about their origin. But the user will not find such different estimations for bilateral trade flows. What are the different estimations for total trade? Why do we not have the same for bilateral trade?

- *Different estimations of total trade*

On the RICardo website, the total trade of a country corresponds to its trade with the partner 'world'. There are up to three different estimations for the total trade of a country. 'World as reported' is the line "total" in original trade statistics tables (primary sources); 'World sum partner' equals the sum of trade partners flows available in the database; 'World estimated' are total trade data collected from indirect sources. It may be the re-estimation of a country's total trade when original data are of bad quality (this is for example the case for Belgium and the Netherlands whose trade statistics include transit trade), or it can be total trade data of countries for which we have no primary source (in this case, we have relied upon national publications of historical statistics or Mitchell's *International Historical Statistics*). From these three estimations, we reconstruct the most complete and of highest quality total trade series, the 'World best guess' series. It is based on the following priority rule: when available 'World estimated' is chosen, then 'World as reported', then 'World sum partner'. 'RIC World best guess' series may be compared on the site to 'World Federico-Tena' series.¹

- *Why do multiple estimations exist for total trade?*

For each reporting entity/year, 'World sum partner' adds up trade flows with all the partners available in the database. This sum should, but does not when the list of partners is incomplete, coincide with 'World as reported'. 'World estimated' does not equal 'World as reported' since this is a re-estimation of a country's original total trade data. 'World Federico-Tena' can differ from 'RIC World best guess' since the data are adjusted and the number of countries is more homogeneous.

- *The adjustment procedure in 'World Federico-Tena' series is difficult to apply to bilateral trade series*

In order to provide consistent series of total trade over a long period of time, Federico and Tena-Junguito make adjustments in the data collected from original sources. These adjustments include: estimation of special trade (when not available in official tables); conversion into current prices for countries/periods when prices of exports and imports were fixed and not revised annually (UK import prices before 1853, French trade prices before 1847, German trade prices before 1881, trade prices in many Latin American countries); exclusion of trade in bullion when it is not categorized as merchandise trade; conversion of trade flows on a c.i.f. (for imports) /f.o.b. (for exports) basis; conversion of fiscal years into calendar years. Federico-Tena series are expressed in gold dollars.

¹ Federico & Tena-Junguito (2016).

Such re-estimation of the value of bilateral trade flows, although desirable, would require a tremendous amount of work. A second-best solution is to establish a ranking of trade statistics according to their quality.

- *Series at constant prices*

The Federico-Tena total trade database provides a series of countries' total trade at constant prices. There are three sources/methods to estimate price indexes to adjust for inflation: existing estimations of price indexes for specific countries; Federico-Tena estimations of price indexes (Laspeyres or Fischer indices) based on London market prices adjusted for freight for a panel of products; a proxy estimation that uses a combination of price series of other countries.²

The RICardo and Federico-Tena databases have two different approaches. RICardo's primary goal is to provide an exhaustive collection of bilateral trade data. It lets the user decide whether to apply an adjustment to convert current values into constant values and which method of adjustment he will choose. Federico and Tena are primarily interested in estimating consistent long-term series and have focused on total trade data. The comparison of RIC 'World best guess' and 'World Federico-Tena' on the RICardo website makes it possible to estimate the divergence between adjusted and non-adjusted series of total trade and to establish a ranking of the trade statistics quality of the world's countries.

CURRENCY CONVERSION

When working with multilateral trade data, it is useful to convert trade values to one common currency. This requires us to compile exchange rates figures for a great number of countries. Exchange rate data have their own set of problems. One problem is the lack of a comprehensive exchange rate database. This meant having to compile exchange rate data for the currencies in which trade values are reported and deciding upon the highest quality exchange rate data.

- *Exchange rate databases*

Historical trade data reported in local currencies are frequently converted to US dollars (Barbieri, Federico-Tena) or to British pounds (RICardo). The *Barbieri* and *RICardo* bilateral trade databases use current exchange rates. The sterling was a reference currency until WWII; US dollars are used by many international organizations after the 1950s. Converting a historical multilateral trade database into a single unit of account requires the use of a historical exchange rate database. We can refer to existing exchange rate databases or decide to reconstruct our own series. Available historical databases are Gurr's Polity II base³, Global Financial Data⁴, and MOxLAD database⁵ (for Latin American countries). These compilations of exchange rate series are ready-to-use but they do not give detailed information on their original sources, do not answer challenging issues, and are limited in their spatial or temporal coverage. Creating his/her own exchange rate database becomes inescapable when the trade database is more exhaustive and includes a lot of "minor" currencies over a long period. In this undertaking, we each have had to look for many sources and have not always found the

² See Federico & Tena-Junguito (2016), 16-20.

³ Gurr T. R. et al. (1989).

⁴ https://www.globalfinancialdata.com/Databases/Data.html#Exchange_Rates_to_1383

⁵ <http://www.lac.ox.ac.uk/moxlad-database>

same. Barbieri's exchange rate database⁶ refers to: Gurr's database, *Statesman's Yearbook*, Bidwell (1970), Gowa and Hicks (2015), Officer (2015), and the United Nations (1962). Federico and Tena's main references are: Global Financial Data, League of Nations' *Review of trade*, *Statistisches Jahrbuch für das Deutsche Reich*, Denzel (2010). In the RICardo Project, we have relied upon: the *Währungen der Welt* collection, Denzel (2010), the League of Nations' *Review of trade*, *Statesman's Yearbook*, and the MOxLAD database.

- *Problems with exchange rates*

Some exchange rate data problems have not been resolved. For example, there are cases where multiple currencies exist for a given country. We have each had to face the Latin American currencies' imbroglia where we have to steer a course between peso, peso fuerte, gold peso, paper peso, and current peso. We also found cases, particularly for Latin American states, where the values of import and export flows are reported in two different currencies, such as gold pesos and silver pesos. This requires separate exchange rates for converting export and import values. Besides, there is the problem of exchange rate variability over the time span of reported trade flows (e.g., exchange rates vary within a given year of annual trade data reports). Exchange rates can be particularly volatile during periods of crisis or uncertainty. There are two historical periods in the last two centuries when currency devaluation was so dramatic that current exchange rates could not be used: the American Civil War (1861-1865) led to the plunge of the Confederate dollar and the explosion of trade values expressed in current dollars; likewise, a spike in trade values expressed in current pounds exists after the depreciation of the British currency after World War I (1915-1923). Gold parity rather than market exchange rates has to be used over these turbulent periods. Nevertheless, other causes and cases of exchange rate fluctuations exist that translate into trade values fluctuations through the process of conversion. The researcher may be tempted to smooth exchange rate values in order to get more regular series. Barbieri reports the example of the *Polity II* exchange rates that have been revised to account for dramatic discontinuities in the time trends of individual exchange rates.⁷

Both Barbieri and RICardo support the option of not smoothing, interpolating or extrapolating the data. They consider that dramatic variations, if not the result of errors, reflect historical reality that should be investigated by scholars. Nevertheless, the COW Trade Data Version 4.0 includes estimates of smoothed data for those scholars who wish to use the data set for pooled cross-sectional time series analysis.⁸

MISSING TRADE

The 'missing trade' issue is one of the most challenging problems we confront. It requires further study because we are uncertain how often data are missing due to lack of trade versus lack of reporting. Government agencies that report trade figures often lack transparency in the techniques used to collect and construct trade reports. Two issues can be distinguished: the definition and the solution of the problem.

- *What are the problems with missing data?*

⁶ Barbieri et al. (2008); Barbieri & Keshk (2016).

⁷ Barbieri et al. (2008), 15-16.

⁸ Barbieri & Keshk (2016), 9.

Trade databases' builders are all confronted with the problem of missing data, particularly when they get into disaggregated data by country or by product. A first reaction is to impute a zero value to missing data. This approach is not ideal, particularly when one state reports a trading relationship and the other does not. Missing data can lead to inconsistencies in trade series and, to inadequacies between total trade data and the sum of bilateral trade data. It raises specific problems for trade gravity models and can have implications for the conduct of trade policies. It is thus an issue that cannot be ignored.

- *Are bad data better than no data?*

The problem of missing data is linked to the problem of reliability of trade data. Much has been written about this subject for more than a century, and one of the most famous defender of skepticism regarding the quality of bilateral trade data was Morgenstern (1963). However, we can read in the foreword of the French edition of his book this balanced judgment made by the French academician Guitton: “*Dès la première édition de 1950, j’avais été frappé par l’immense portée des réflexions que suscitait cet ouvrage. J’enseignais alors la Statistique. Le problème des erreurs me paraissait fondamental à faire comprendre, non pas pour dénigrer la Statistique, comme cela est arrivé si souvent d’une manière naïve et même un peu enfantine : ne pas utiliser des données statistiques sous prétexte qu’elles ne sont pas exactes ! C’est pourquoi je me suis employé dans ma petite sphère à habituer les esprits à travailler dans l’imperfection des saisies chiffrées des phénomènes, pour en tirer cependant le maximum de vérité. La philosophie des erreurs commande l’intelligence même de la Statistique.*”⁹

In this field, it is interesting for the economic historian to turn to the statistician. From his daily experience, the statistician knows that: no statistic is perfect; it would be illusory to think that reality can be grasped without any error; it is not unusual to replace missing data with estimated data.¹⁰ As explained in the codebook¹¹, the Barbieri and COW Trade Data Set have taken the option of substituting a missing bilateral import value with the exporting state’s report for the same flow (mirror flow). It also relies upon the exporter’s report when the importing state reports zero trade and the exporter reports trade values. A different approach has been adopted in RICardo which, unlike Barbieri, focuses on the pre-WWII period where it can be assumed that sources of divergence in mirror flows were more numerous. In RICardo, we recommend that the user takes advantage of the comprehensive character of the base to compare data and ultimately select the more reliable figures.¹²

Once the decision to impute missing trade values has been taken, different techniques can be used: mirror statistics (after adjusting for known existing bias, e.g., f.o.b./c.i.f.); interpolation (between two known data) or extrapolation (before or after known data to complete a time series).

⁹ Henri Guitton in Morgenstern (1972), foreword. “*As soon as this book was published in 1950, I was immediately convinced that the kind of thinking it entailed had huge implications. The issue of errors seemed to me crucial to understand, and not merely in order to denigrate Statistics, as so many did so often in rather naïve, even childish way —as if one should not use statistical data on the grounds that they are not perfectly correct! This is why I strived, within my own small sphere of activity, towards acclimatizing the minds of others to the need to work with the imperfections of quantified manifestations of phenomena so as to still derive from them as much truth as possible. The philosophy of errors is the essential vantage point for the very comprehension of Statistics.*” We thank Pierre Gervais for his help in translating the Guitton’s quotation.

¹⁰ As is the case in the IMF bilateral trade statistics. See IMF (1993), 9-10.

¹¹ Barbieri & Keshk (2016), 4-8.

¹² Dedinger & Girard (2017), 36-38.

The missing trade issue remains a challenge that requires the expertise of both the economic historian, for his knowledge of statistical documents and historical context, and the statistician, for his knowledge in techniques of imputation.

TRADE VALUATION

Beginning in the seventeenth century the systematic compilation of trade statistics was slowly organized by state agencies.¹³ Trade data were registered and published either in quantity¹⁴ or, more usually, in value. The valuation of exported and imported merchandises raises a number of issues that have been thoroughly analyzed in the past, particularly by Petruzzelli (1946) and Allen & Ely (1953).¹⁵ Yet, the availability of exhaustive and disaggregated trade databases opens new opportunities for research on this complex matter. For a better understanding of the issue, we have focused on a case study taken from the *Montesquieu* database.

- *How are set trade values?*

The *Montesquieu* database documents French bilateral trade statistics by products between 1836 and 1938 collected from the French publication “*Tableau général du commerce de la France*”. From 1827 (first year of publication) to 1846, trade values were set by using official prices (i.e. fixed prices). Beginning in 1847, actual prices (i.e. variable prices) were set annually by a *Commission des valeurs en douane* which centralized price information given by the *Chambres de commerce* and *Chambres consultatives des arts et manufactures*. The two series, actual and official values, are published over 1837-1846 showing a divergence. Nevertheless, can we argue that French trade values are more reliable after 1847? Comparing French bilateral trade flows computed in French publications with the same flows computed in partners’ publications – i.e. mirror flows – provides a lot of information on the relative quality of trade values. Let’s turn to the example of Franco-Belgian trade in 1910.

- *A case study to better understand trade valuation: Franco-Belgian trade in 1910*

Bilateral flow	<i>Belgian source</i> (1000 £)	<i>French source</i> (1000 £)	<i>Mirror flows discrepancy*</i>
Exp _{Belg-Fr} vs Imp _{Fr-Belg}	26191	18621	0.29 (-0.41)
Imp _{Belg-Fr} vs Exp _{Fr-Belg}	29253	39782	-0.36 (0.26)
[Exp _{Belg-Fr} /Imp _{Belg-Fr}] vs [Imp _{Fr-Belg} /Exp _{Fr-Belg}]	0.9	0.5	

* For the computation of mirror flows, see the bilateral view on the RICardo website. The first result is based on the Belgian source, the one in brackets on the French source.

Source: <http://ricardo.medialab.sciences-po.fr>

There is a relatively great divergence in the Franco-Belgian trade values registered by the two states that cannot be attributed to transport costs alone (imports are valued c.i.f., including fret and insurance, whereas exports are valued f.o.b.). By focusing on a sample of textile products – considered as homogeneous and comparable – it appears that French statistics overvalued the quantity and value of textile exports to Belgium compared with Belgian registrations for equivalent imports.

¹³ See Charles & Daudin, eds. (2015).

¹⁴ Cf. the case of the German Zollverein, Dedinger (2015).

¹⁵ For a specific example, see Charles & Daudin (2011).

	Imp _{Belg-Fr} / Belgian source			Exp _{Fr-Belg} / French Source			FUV/ BUV
	Quantity (tonne)	Value (1000 Frs)	Unit value (BUV)	Quantity (tonne)	Value (1000 Frs)	Unit value (FUV)	
Woollen fabrics	781	6465	8.28	1751.5	20877	11.92	1.44
Cotton fabrics	1078	9243	8.57	3016.9	31554	10.46	1.22
Silk fabrics	178	7690	43.20	236.2	15713	66.52	1.54

Sources: *Tableau général du commerce de la France*; Statistique de la Belgique – *Tableau général du commerce*.

Many questions arise from this table. Why such a difference in quantities and values between these two neighbor countries? Did national *Commissions* use different rules to set import and export prices? Did France/Belgium apply the same prices to all trade partners and to all national provinces or regions? Do we really compare the same flows of products? For example, values are supposed to be of special trade for the two countries, but it is known that Belgian special trade can include transit trade.¹⁶ Is there a constant bias through time? To some extent, valuation differences are expected as trade statistics report most of the time national prices rather than bilateral ones. It might be that Belgian imports of cotton fabrics from the United Kingdom were less expensive than the imports from France, or that French exports to Italy were more expensive than French exports to Belgium.

These are all genuine research issues to which we would like to draw the attention of the user. And perhaps rouse researcher's new interest in this matter.

MIRROR FLOWS

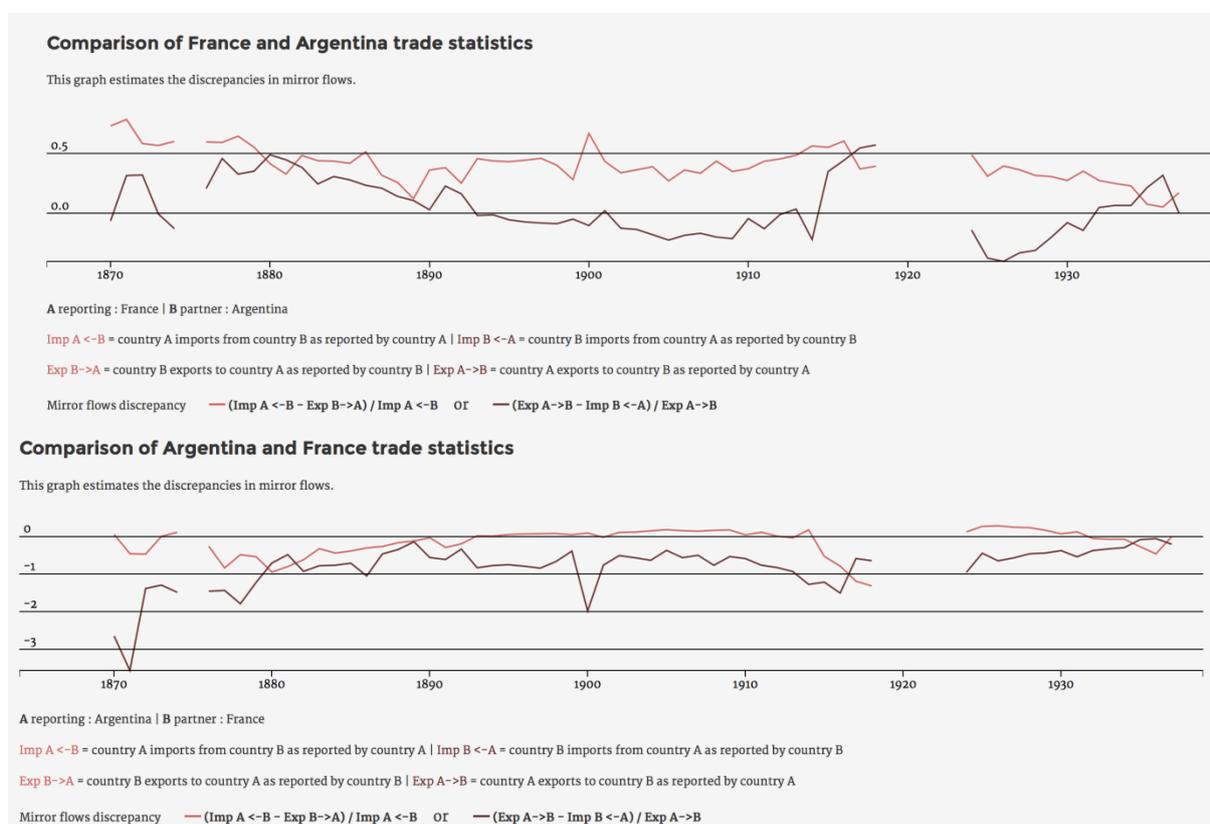
A rare feature of trade statistics is that each bilateral trade flow is collected and reported twice, once by the exporting country and once by the importing country. Whereas this is theoretically the same trade flow, we have already noticed that the two data reports are never the same value. There are many causes for the more or less important differences between two similar reports. The most common ones are: varying procedures for recording flows by country of origin or destination, different methods for estimating trade flows (c.i.f. vs. f.o.b. estimates, trade types), or the choice of the exchange rate. To what could be added deliberate misreporting.¹⁷ In this point, we would like to focus on the RICardo website that provides a computation and visualization of mirror flows in the bilateral view. The committee was asked to give its opinion on alternative utilizations of the RICardo's mirror flows.

¹⁶ 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. Nevertheless, in the definition given by the League of Nations, transit trade is explicitly included in general trade (Société des Nations 1928, 10). 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 *Barbieri*, *Federico-Tena* and *RICardo* databases have favored special over general trade when both types of information were available.

¹⁷ See for example Kellenberg & Levinson (2016).

- The RICardo bilateral view

The bilateral view allows the user to chart bilateral trade between two countries reporting trade statistics over a selected period of time. It also provides a view of the mirror flows' discrepancy. There are four values for each country-pair. Exp_{A-B} (exports from A to B as reported by A) can be compared to Imp_{B-A} (Imports of B from A as reported by B) – and respectively Imp_{A-B} to Exp_{B-A} . For the calculation of the mirror flows gap, we have relied upon Morgenstern¹⁸ (see the computation formula in the graphs below). The ratio should reflect only c.i.f. trade costs and be relatively constant through time but, as the bilateral view makes it possible to visualize, it greatly varies in practice. The graphs below illustrate the mirror flows problematic by comparing bilateral trade between France and Argentina on the basis of French and Argentinian trade statistics documents. From a French point of view, French values tend to overestimate trade with Argentina in a margin of [0;0.5]; from the Argentinian point of view, trade with France is underestimated in a much greater margin, on average [-1.5;0].¹⁹



- Alternative uses of the RICardo's mirror flows?

The 'bilateral view' has been created to complement the other views of the RICardo website. Unlike the 'country view' and the 'world view', it is not intended to give synthetic or aggregated visions of historical trade relationships. It gets back to the origin of the data and invites to further investigation on this specific issue. Besides, it wants to make the user aware of the mirror flows problematic. But it can also have the side effect of confusing the uninformed user. The RICardo team has invited the other members of the committee in a

¹⁸ Morgenstern (1963), 170.

¹⁹ Spikes and dips in the curve can actually be errors. The bilateral view also provides a useful means of detecting errors in the database.

reflection on the bilateral view and the potential uses of the RICardo's mirror flows. It is very thankful for all the suggestions made. A first question was to specify our objective, what we want to do with the mirror flows. If the purpose is to estimate the quality of trade statistics, it is more interesting to get a picture of all the mirror flows of one country. For example, by looking at France's – and Argentina's – all bilateral flows, it would tell us if France systematically tended to overestimate – Argentina to underestimate – the value of its trade flows. The mirror flows of the database could therefore be used to graduate reporting countries according to the reliability of their trade statistics. This rating could then serve to fill in missing data in the RICardo database in a "scientific" way and to provide the user with a complete and ready-to-use database guaranteed by expert evaluation.

New reflection is required on the RICardo bilateral view and the exploitation of mirror flows. It would be interesting to have a feedback from the users of the website.

ENTITIES

With this point we get into the second set of issues discussed by the committee. It relates to the exploitation of the RICardo database with digital technology, which is the most innovative part of the project. Economic historians, developers and designers have collaborated to create the RICardo website. As it keeps on improving the site, the RICardo team submitted to the committee some delicate questions it has to face. Let us first recall that RICardo is a bilateral trade database which serves to study the distribution of an entity's trade by partners. The term 'entity' has been privileged because there are many cases in historical trade statistics where the trading entity does not correspond to an internationally recognized state but describes a locality, an area or a group of states. "Entities" are the basic information to analyze and to compare the distribution of trade. There are 'reporting' and 'partner' entities, the list of partner entities being much more longer and heterogeneous. A great part of the project has been devoted to the construction of a RICardo entity table in order to standardize the names of the entities and to establish a system of codification of these entities. Much more could be carried out, such as the creation of aggregated entities or the introduction of a political codification for the entities. But this raises problems due to the non-homogeneity of entities that need to be clarified. After a presentation of the structure of the RICardo entity table, we examine these problems and the comments of the committee.

- *The RICardo entity table*

The RICardo entity table is made up of four columns. 'RICname' is the standardized name of the entity. The COW entity table has been used as much as possible, more precisely the document entitled 'Entities.pdf'.²⁰ Each RICname is associated to a type. As can be seen in the table, five types of entities are distinguished. The 'country' type corresponds to a COW name. 'Group' is a group of entities of the 'country' type. The other RICnames have been created. The last two columns specify the continent to which the entity belongs, and the COW code for the 'country' type. In the first version of the database (June 2016), the 5277 original names have thus been reduced to a number of 1598 RICnames.

²⁰ More precisely the document entitled '*Entities.pdf*'. See *Correlates of War Project. Colonial Contiguity Data, 1816-2016. Version 3.1. Online*, <http://correlatesofwar.org>

RICname	type	continent	COW_code
Aden & Algeria & Egypt/United Arab Republic & Iran (Persia) & Asian Turkey	group	World	
Adriatic harbours	city/part_of	Adriatic	
Afghanistan	country	Asia	700
Ancona	country	Europe	3271
Andaman and Nicobar Is.	country	Asia	7501
Andorra	country	Europe	232
Angola	country	Africa	540
Angola & Kabinda (Portuguese Congo)	group	Africa	
Angola & Portuguese West Africa	group	Africa	
Angra Pequena	city/part_of	Africa	
Anjouan	country	Africa	5813
Annaba (Bône)	city/part_of	Africa	
Belgium & Luxemburg	group	Europe	
Belize (British Honduras)	country	America	80
Belize (British Honduras) & British Virgin Is.	group	America	
Black Sea & Levant	geographical_area	World	
British Africa	colonial_area	Africa	
Central Africa	geographical_area	Africa	
Central America	geographical_area	America	
Danish America	colonial_area	America	
Danish Colonies	colonial_area	World	
Danish Colonies & Denmark	group	World	
Dominica	country	America	54
Dominican Republic (Santo Domingo)	country	America	42
Dominican Republic (Santo Domingo) & Haiti	group	America	
Dunkirk	city/part_of	Europe	
Dutch America	colonial_area	America	
Dutch Asia	colonial_area	Asia	

- *RICentities' flaws*

To analyze the distribution of a country's trade at a given time or its growth over time, we should theoretically use homogeneous entities, that is, entities whose definition is constant over time. But we know that this is not the case in reality, that the definition of an entity (political, geographical, statistical) is not constant through time. There are different kinds of problems in the RICardo database.

Border changes issue: This is a recurring issue that any user has in mind. Federico & Tena (2016) estimate series of total trade at constant borders (as of 1913) for 149 polities.²¹ This does not seem possible with a bilateral trade database. The question is more how to inform the user of this border issue on the website.

Disaggregation of groups of entities: This problem is not often mentioned. It appears in the RICardo database because data have been collected as they appear in the original documents. In these documents, many partner entities are of the 'group' type. To illustrate the problem, let us see the example of bilateral trade between France and China over 1827-1938. In French trade statistics China (People's Republic of China) is sometimes aggregated with other entities making it problematic to reconstruct a homogeneous historical series.

'China' partner in France's trade statistics 1827-1938:
- China & Cochin China 1827-1836
- China & Cochin China & Oceania & Philippines 1837-1840
- China & Cochin China & Oceania 1841-1866
- China 1867-1938

21

Federico & Tena (2016), 20-26.

No political dimension: as of now, there is no codification of the RICentities allowing the identification of dependent and sovereign states in time. It seems that such a political classification would be greatly recommended to reconstruct colonial aggregates, or customs unions, or federal states and provide the user with aggregated trade series.

- *Is the creation of a ‘perfect’ database a practicable option?*

When examined in detail the idea of overcoming all the difficulties linked to the definition of entities becomes illusionary. Many points have been raised by the members of the committee.

What is the territorial definition of an entity? For example, do we consider that French borders are limited to metropolitan France or does it also include overseas territories? What about Alsace-Lorraine? Do all trade statistics sources refer to the same territory for each entity?

The difficulty of creating homogeneous aggregated trade series in time: in the above example, how could we recalculate a homogeneous French trade series including Alsace-Lorraine, the Martinique and the Guadeloupe over the whole RIC period? Another issue is when a reporting entity becomes part of a sovereign state, as is the case with federated states. With the creation of a federation (for instance Australia, Canada, Italy, South Africa...), “national” trade statistics for federated states are generally no longer published.

How to define an aggregated entity? Suppose we want to create the aggregate “British empire”. What would it include? How to choose a specific date to define it? Such examples are infinite.

The committee advises the RIC team to avoid creating too much complexity that would lose rather than help the user. There are two main recommendations.

Create homogeneous aggregates for some major countries: relying on one trade source, for example the French *Tableaux du commerce*, homogeneous trade aggregates series can be provided to the user, such as French trade with French colonies, Central America, etc.

Let the user create his/her own aggregated series: given the complexity of the issue, the best option is to invite the user to be active and creative. And to provide him with the means of constructing aggregated/disaggregated series by putting on the website some documentation. Using the example of the COW entity table (cf. below), the RIC entity table could be supplemented with information on the political status of RICentities through time and be made available in an Excel file or PDF document.

Entity Number	Name	Begin Year	End Year	Ending Political Status
3	Alaska	1816	1867	Became colony of 365
3	Alaska	1867	1959	Became colony of 2
3	Alaska	1959	1993	Became part of 2
4	Hawaii	1898	1960	Became colony of 2
4	Hawaii	1960	1993	Became part of 2
5	Virgin Islands	1816	1917	Became colony of 390
5	Virgin Islands	1917	1993	Became colony of 2
6	Puerto Rico	1816	1821	Became part of 1070
6	Puerto Rico	1821	1898	Became colony of 230
6	Puerto Rico	1898	1952	Became colony of 2
7	Texas	1816	1836	Became part of 70
7	Texas	1845	1861	Became part of 2
7	Texas	1861	1865	Became part of 1002
7	Texas	1865	1993	Became part of 2
10	Greenland	1816	1953	Became colony of 390
10	Greenland	1941	1945	Occupied by 2
11	Faeroe Is.	1816	1947	Became colony of 390
11	Faeroe Is.	1940	1945	Occupied by 200
20	Canada	1816	1841	Became part of 1025

The discussion led us to the conclusion that all the problems cannot be resolved. Some pedagogical work has to be made for the user eager to properly use the database by providing him/her with sufficient information.

FUTURE PROSPECTS

The workshop concluded on future prospects regarding the RICardo database and website. A second version of the database is currently underway. About 100 000 new data points extracted from primary sources (national customs returns) found at the Insee (*Institut national de la statistique et des études économiques*) will be added to the database. This new version will be ready for the official launching of the RICardo website by the end of 2017. The RIC team is already thinking of new innovations for the project – conditional upon the finding of new funds – which it submitted to the committee.

New features on the website

As an extension of the discussion point ‘Entities’, a new feature should be created to allow for the creation and visualization of aggregated entities. Besides, the user shall be given the opportunity to select not just reporting entities but also partner entities. He/she may want to know for instance which countries report trade with ‘China (People’s Republic of China)’ over the period. Another objective is to exploit the potential of the exchange rate database by creating a new view. And we would like to create annotations giving informations on structural breaks in the series due to historical events (for example, a user pointed to the dramatic increase in Danish exports in 1850 due to the momentary integration of the Duchies of Schleswig and Holstein in the Kingdom of Denmark).

Updating the database

As RICardo is an evolutionary database, it is intended to make new updates in the future either by collecting new sources or by merging with other databases. As we have experienced in working on the second version of RICardo, the update work can be a very long and tedious process. It has to be improved in two ways: by introducing a tool to edit the data directly in the structured database linked to the versioning system and by offering a system to check and provide feedback about potential errors available not only for the administrator but also for the user.

User interface

The RICardo website provides the user only with exploration tools. We could imagine a deeper interaction by providing editing features which would allow users to write historical or economic analysis or comments directly on the website. Yet the site is also meant to serve as a tool for disseminating knowledge and uses of historical trade statistics in the public. It is too early to judge of the potential impact of the website. Is it attractive only to a restricted group of scholars? Can it extend to a wider audience of students or teachers? The kind of interface we introduce will be dependent on the reaction of the public. In a first time, we will open a blog on the Hypotheses platform to disseminate information on the RICardo project and its uses (<https://ricardo.hypotheses.org/>).

CONCLUSION

This brainstorming workshop has yielded the most positive results. The RICardo team is very grateful to all participants for their work, comments and suggestions that, we hope, can be helpful for potential users of historical trade statistics and the RICardo website. Users' questions and reactions to any of the discussion points are welcome.

REFERENCES

- Allen R. G. D. & J. E. Ely (1953). *International Trade Statistics*. New York; London.
- Barbieri K., O. M. G. Keshk, & B. M. Pollins (2008). *Correlates of War Project Trade Data Set Codebook. Version 2.01*.
- Barbieri K. & O. M. G. Keshk (2016). *Correlates of War Project Trade Data Set Codebook. Version 4.0*. (<http://correlatesofwar.org>)
- Bidwell R. L. (1970). *Currency conversion tables: a hundred years of changes*. London: Collins.
- Bécuwe S., B. Blancheton, & K. Onfroy (2015). Données du commerce extérieur de la France entre 1836 et 1938 – Base Montesquieu. *Cahiers du GREThA*, n° 2015-35. (<http://ideas.repec.org/p/grt/wpegrt/2015-35.html>)
- Charles L. & G. Daudin, eds. (2015). Eighteen-Century International Trade Statistics. Sources and Methods. *Revue de l'OFCE*, 140.
- Charles L. & G. Daudin (2011). “La collecte du chiffre au XVIIIe siècle : le Bureau de la balance du commerce et la production des données sur le commerce extérieur de la France ». *Revue d'histoire moderne et contemporaine*, 58 (1), 128-155.
- Dedinger B. (2015). “Trade Statistics of the Zollverein, 1834-1871”, *Revue de l'OFCE*, 140, 67-85.
- Dedinger B. & P. Girard (2017). “Exploring trade globalization in the long run: the RICardo project”, *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 50(1), 30-48.
- Denzel, M. (2010). *Handbook of World Exchanges Rates, 1590-1914*. Surrey: Ashgate Publishing, Ltd.
- Escaith, H. (2015). “Past and present issues in trade statistics”. *Revue de l'OFCE*, 140, 39–51.
- Federico G. & A. Tena-Junguito (2016). “World trade, 1800-1938: new data-set”. *EHES Working Papers in Economic History*, n° 93.
- Findlay R. & K. H. O'Rourke (2007). *Power and Plenty: Trade, War, and the World Economy in the Second Millennium*, Princeton New Jersey, Princeton University Press.

- 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.
- Gurr T. R., K. Jagers, & W. H. Moore (1989). *Polity II Handbook*. Boulder: University of Colorado, mimeo.
- Gowa J. & R. Hicks (2015). “Politics, Institutions, and Trade between the Wars: New Data and Conventional Wisdom”, *International Organization*, 67(1), 439-467.
- International Monetary Fund (1993). *A guide to direction of trade statistics*. Washington DC.
- Kellenberg D. & A. Levinson (2016). *Misreporting Trade: Tariff Evasion, Corruption, and Auditing Standards*. NBER Working Paper, n° 22593.
- 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.
- Officer L. H. (2017) “Exchange Rates between the United States Dollar and Forty-One Currencies”. *MeasuringWorth* (<https://www.measuringworth.com/exchangeglobal/>)
- O'Rourke K. H., ed. (2005). *The International Trading System, Globalization and History*, 2 volumes, Cheltenham: Edward Elgar.
- Petruzzelli N. M. (1946). *Some Technical Aspects of Foreign Trade Statistics with Special Reference to Valuation*. Washington D.C.: Catholic University of America Press.
- 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. Lateinamerikanische 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.

United Nations (1962). *International Trade Statistics 1900-1960*.
(<https://unstats.un.org/unsd/trade/data/tables.asp#historical>)

United Nations (1998). *International merchandise trade statistics: Concepts and definitions*.
New York: United Nations.