Exploring Practices of Conflict Data Production, Analysis, Dissemination, and Practitioner Reception: Methodological Framework and Preliminary Findings

Grey Anderson, Louise Beaumais, Louise Carnapete, Iris Lambert, Sami Makki, Frédéric Ramel, Eric Sangar

To cite this version:

HAL Id: hal-03218342

https://sciencespo.hal.science/hal-03218342

Preprint submitted on 5 May 2021

THE DATAWAR PROJECT

Exploring Practices of Conflict Data Production, Analysis, Dissemination, and Practitioner Reception: Methodological Framework and Preliminary Findings

Working paper 2021/1

Contributors

Grey Anderson (Ecole polytechnique / LinX)
Louise Beaumais (Sciences Po / CERI)
Louise Carnapete (University of Bath, Humboldt-Universität zu Berlin, Università degli Studi di Siena)
Iris Lambert (Sciences Po / CERI)
Thomas Lindemann (Ecole polytechnique / LinX, Head of WP1[1])
Sami Makki (Sciences Po Lille / CERAPS, University of Lille)
Frédéric Ramel (Sciences Po / CERI, Head of WP2[2])
Eric Sangar (Sciences Po Lille / CERAPS, University of Lille, Head of WP3[3])

Abstract

How are representations of violence influenced by the 'agency of data', in other words the social practices of data collection, analysis, dissemination, and practitioner reception? The DATAWAR project builds on the hypothesis that scientific output in quantitative conflict studies is driven less by theoretical innovation than by the 'politics of data': the availability, reputation, and mathematical malleability of numerical observations of conflict. We anticipate that the perceptions of conflict developed by practitioners who employ quantitative methods and sources are prone to distortion as a result of the nature of the available data, the type of mathematical models used to analyse and potentially 'predict' conflict, and reliance on a selective subset of theoretical approaches. DATAWAR will carry out the first systematic investigation of scientific practices in the field of quantitative conflict studies as well as the impact of these practices on practitioners' vision of war, covering the full lifecycle of conflict data, from collection and analysis to their use and dissemination by military and diplomatic institutions, humanitarian organisations, and the media. The unique, cross-actor and cross-national perspective of DATAWAR aims to improve our understanding of the interactions between scholarly and applied uses of conflict data, beyond the established divide separating 'data pessimists' and 'data optimists'.

---

[1] Work Package 1: Practices of the production of quantitative conflict knowledge.
[2] Work Package 2: Analysis of impact on practitioners' perceptions and representations of conflict.
[3] Work Package 3: Dissemination of results, training and advice for civil and military practitioners.

Introduction

The DATAWAR project seeks to explore the following research question: how do social practices of data collection and analysis in quantitative conflict studies influence practitioners' representations of armed conflict?

DATAWAR does not aim to exploit data in order to develop and test yet another large-N statistical model to 'explain' or forecast international violence. Instead, the project examines how academic scholars produce and analyse quantitative data on armed conflict and how these practices shape perceptions and interpretations on the part of professionals in the field of conflict management as well as the media. This ambition responds to widespread calls for greater reflexivity with respect to often-overlooked biases and potential side-effects of data-driven and algorithm-based analysis of human behaviour. In an age when more and more private and public actors are turning to 'big data' to understand and even predict political conflict and violence, there is a surprising and worrying knowledge gap between day-to-day practices of scientific data collection and analysis, on the one hand, and practitioners' perceptions and normative conclusions concerning the causes, dynamics, and management of armed conflict, on the other.

To this end, DATAWAR investigates both scientific practices in the field of quantitative conflict studies and the impact of these practices on practitioners' vision of war, covering the full lifecycle of quantitative conflict data, from collection and analysis to their use and dissemination by military and diplomatic institutions, humanitarian organisations, and the media.
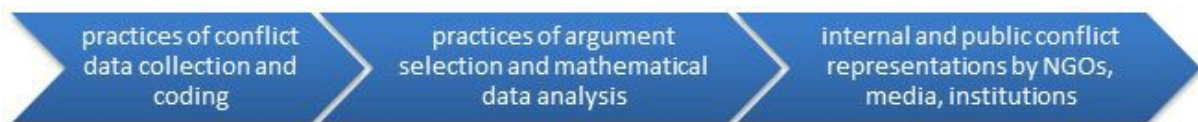
We conceive of 'impact' here not in terms of a narrow causal link, by which a given scientific discourse would directly 'produce' a specific type of perception, but rather as all of the ways in which actors receive, filter, and interpret the output of quantitative conflict analysis. Thus, contrary to a strict positivist perspective that takes data as neutral, we focus on how practices governing the very definition of armed conflicts, the collection and coding of quantitative data, and the mathematical testing of 'dependent' and 'independent variables' affect the perception of armed conflict, its causes and dynamics. To give an example: whether armed conflict is defined according to a threshold of 25 as opposed to 1,000 battle-related deaths will play an important part in determining the perceived frequency of war in international relations and contribute to a more or less pessimistic outlook on the prospect for future conflagrations.[4]

Drawing on insights from Bruno Latour's sociology of science (Latour, 1987; Mackenzie, 1991) and Theodore M. Porter's critique of the politics of quantification (Porter, 1995), this project explores the hypothesis that scientific output in quantitative conflict studies is driven less by theoretical innovation than by the 'politics of data'; that is to say, the availability, reputation, and mathematical malleability of numerical observations of conflict. One consequence of the 'agency of data' in quantitative conflict studies may be to prioritize nomological theoretical approaches with little or no attention to contextual, hermeneutic factors. Although quantitative methods are not intrinsically incompatible with interpretative

---

[4] An intriguing illustration of resulting differences in the visual representation of international violence can be found in Dieckhoff et al. (2016, p. 256).

or critical approaches (Baillat, Emprin, & Ramel, 2016; Balzacq, 2014), we expect to find a preference for theoretical concepts that can be easily tested using available large-N datasets, and whose empirical indicators do not require in-depth research (Lindemann, 2016, pp. 44-48). As a result, we anticipate that the perceptions of practitioners who rely on quantitative conflict research are likely to be biased by the nature of the available data, the type of mathematical models employed to analyse and potentially 'predict' conflict, and reliance on a selective subset of theoretical approaches.

DATAWAR's Methodological Framework

Why is this project relevant?

Many of the most influential accounts of the causes and dynamics of armed conflict have benefited from quantitative research, as reflected in the contents of high-impact factor academic journals such as the *Journal of Peace Research* or the *Journal of Conflict Resolution* (Gleditsch, Metternich, & Ruggeri, 2014). Indeed, key results of conflict studies that are known among practitioners and the media – such as the 'Democratic Peace' thesis – have become established only after thorough quantitative testing and refinement, while only a small percentage of the most frequently cited articles on the subject do not make use of quantitative methods.

The history of quantitative conflict research is closely linked to the emergence of International Relations (IR) as a distinctive social science discipline, whose theoretical and methodological foundations were for the most part first laid in the Anglo-American world. In the face of resistance from pioneers of 'classical' IR such as Hans Morgenthau, who worked largely with historiographic, hermeneutic methods, behaviouralist scholars such as Kenneth Waltz argued in favour of the formulation of general theoretical laws that could be used to explain – and also 'predict' – the behaviour of states and the outbreak of war across an array of spatial and temporal contexts (Kratochwil, 2006). On this view, understanding the phenomenon of war requires comparison between a large number of cases and the specification of variables and correlations so as to test theoretically formulated causal mechanisms.

Construction of the first generation of databases 'measuring' conflict dynamics and potential causal indicators began in earnest during the so-called behaviouralist turn of the 1960s (Dieckhoff, Martin, & Tenenbaum, 2016, p. 251). Since then, work based on large-N datasets (e.g., those produced by the Correlates of War Project (COW) and the Peace Research Institute of Oslo (PRIO)) has helped to bolster and refine a range of propositions, such as the decline of interstate conflict since 1990 (Harbom, Melander, & Wallensteen, 2008) or the role of foreign aid in explaining the success of post-conflict reconstruction (Girod, 2015). In the same way, political institutions, NGOs, and the press have all drawn on insights from quantitative studies to forecast armed conflict and adapt their analytical and normative stances (Colonomos, 2016; De Franco & Meyer, 2011; Meyer, De Franco, & Otto 2019; Ward et al., 2013).

However, it appears that a number of recent theoretical shifts in the study of conflict, such as the 'pragmatic turn' or the even more recent 'emotional turn' (Bauer & Brighi, 2008; Clément & Sangar, 2018), have been less frequently taken up by quantitative conflict analysis than traditional, 'objective' considerations, such as the distribution of military power or the scarcity of economic or ecological resources. In other words, the empirical work done by quantitative scholars may only partially reflect important, ongoing theoretical debates in the study of conflict. Furthermore, to date there has been little

systematic study of the actual research practices that go into the collection, coding, interpretation, and dissemination of quantitative conflict analysis. There is also practically no research available on the ontological and normative impact of quantitative conflict research on practitioners (such as government officials, media actors, or NGO analysts). It may be the case that practitioners' perceptions are distorted as a result of the bias in quantitative scholarship towards 'material', non-ideational causes of conflict. Last but not least, it has been confirmed that the nature of the data in the major quantitative conflict databases differs greatly, with important consequences for the interpretation of the causes of conflict and ensuing normative conclusions (Eberwein & Chojnacki, 2001).

In recent years, an expanding body of research has sought to grapple with the 'politics of numbers' and 'governance by numbers' (Supiot, 2015). This literature mainly focuses on how numbers are used by *governments*, whether to control populations at home or wage war abroad (Fioramonti, 2014; Franz, 2017; Greenhill, 2010; Hansen & Porter, 2012). One important finding within this scholarship, developed by Thierry Balzacq and others, is that in international security quantitative data are typically used for the purposes of persuasion, (de)politicisation, and standardisation (Baele, Balzacq, & Bourbeau, 2017). Quantitative data thereby serve to enhance the 'rationality' of governing and controlling populations – although rival social actors can attempt to thwart its instrumentalisation or utilise it to resist political domination (Bruno, Didier, & Prévieux, 2014). Other recent studies have explored the (mis)uses of quantitative conflict studies in internal bureaucratic struggles over the allocation of organizational resources (Beerli, 2017).

What is missing from this literature, however, is a rigorous examination of the ways in which scientific practices themselves, including the internal logics of data collection and academic publishing, shape how practitioners perceive, interpret, and 'predict' armed conflict. This lack of interest in the practice of data collection is all the more surprising considering that some quantitative scholars of conflict have themselves spoken out to criticize the state of the field. A reviewer for the *Journal of Peace Research*, for example, observes: '*In recent years, I have found myself increasingly frustrated with the quantitative papers I am sent to review, whether by journals or as a conference discussant [...] the typical paper has some subset [...] of the following irritating characteristics: [...] Uses a dataset that has been previously analyzed a thousand or more times; [...] The reported findings are the result of dozens – or more likely hundreds – of alternative formulations of the estimation*' (Schrodt, 2014, p. 287).

We believe that informal and – to the extent that they are generally not disclosed in academic publications – hidden practices of data collection, analysis, and dissemination weigh more heavily than is commonly recognized in the framing and interpretation of armed conflicts. It is a paradoxical situation, in that a group of datasets are taken for granted as authoritative sources and the coin of academic excellence, while the very scholars who exploit them acknowledge the more or less serious limits they pose to understanding the phenomena they putatively canvass. This impression is confirmed by a recent study of quantitative data in the context of peace-keeping operations, which concludes that '*the collection and use of data replicate and are poised to extend the theory-practice divide that exists between researchers who study violence – those working 'on' conflict – and the peacekeepers, peacebuilders, and aid workers who work 'in' the midst of it*' (Fast, 2017, p. 706).

Only a handful of scholars have noted that quantitative datasets often implicitly promote a particular conception of war, for example by suggesting an analogy with disease, with epidemiology as the appropriate method to "cure" conflict (Duursma & Read, 2017; Freedman, 2017). Others have criticized

conflict studies for failing to engage with the insights of quantum theory in natural sciences and the inevitability of non-deterministic processes (Der Derian, 2013). Habitual use of linear mathematical models might likewise be thought to encourage the straightforward deduction of future threats from observed demographic or ecological trends. As a result, the future too often looms as an apocalyptic version of the present, with contemporary menaces deterministically amplified. This form of 'presentism' (Hartog, 2003) seems to be linked to a degree of pessimism and scepticism concerning the possibilities for peaceful change that has been detected in threat assessment and conflict forecasting exercises (Colonomos, 2016).

At the same time, the greater availability of conflict databases and related scholarly output (as a result of open-access initiatives) makes it increasingly likely that practitioners will encounter such scholarship in their daily work routines. The UN, for instance, has developed quantitative systems to survey conflict perceptions and the needs of local populations, although research document 'how UNAMID staff and local populations move often in parallel worlds, and how this distance is being maintained even in encounters deliberately instigated to collect data on violent incidents' (Müller & Bashar, 2017, p. 775). The EU's 2016 'Conflict Early Warning System' contains a 'Global Conflict Risk Index' which identifies a number of material and institutional risk indicators yet does not take account of discursive factors (such as the proliferation of 'hate speech' or demands for recognition (Lindemann, 2011).[5] Moreover, some scholars suggest that practitioners equipped with 'big data' may be inclined to reject theoretical explanations altogether, ushering in 'a new era of empiricism, wherein the volume of data, accompanied by techniques that can reveal their inherent truth, enable data to speak for themselves free of theory. The latter view has gained credence outside of the academy, especially within business circles' (Kitchin, 2014, p. 130).

Against this backdrop, DATAWAR intends to trespass the aforementioned 'theory-practice divide' and scrutinise conflict-data practices across institutional boundaries, taking into account specific organisational cultures, needs, and mission profiles. In so doing, the project situates itself in a context in which the initial enthusiasm for data as a tool for conflict prediction and prevention, summarised by Sheldon Himelfarb's article 'Can Big Data Stop Wars Before They Happen?' (Himelfarb, 2014), has given way to more cautious, realistic discussion. DATAWAR's mixed-methods approach, combining a quantitative examination of the uses of conflict data with a hermeneutic inquiry into how scholars and practitioners incorporate conflict data in their daily routines, will contribute to a nuanced, comparative assessment.

In the following, the methodological framework of each of the project's basic analytical components will be detailed.

Analysing scientific practices of conflict data collection, analysis, and dissemination

As the use of quantitative methods in the study of war has effloresced over the past two decades, so too has criticism of their dominance, associated with ills ranging from the decline of grand theory to mathematised obtuseness and loss of policy-relevance (Mearsheimer & Walt, 2013; Desch, 2019). If the rise of quantification can be dated with reasonable precision, the explanation for its take-off might be

---

[5] See: 201409_factsheet_conflict_earth_warning_en.pdf (europa.eu)

thought similarly straightforward (Li, 2019). The parallel development of computing technology and relevant software systems has transformed the arduous, time-consuming and costly coding and analysis of data, once performed by teams of undergraduate and graduate assistants, and brought advanced statistical tools to the personal computers of individual researchers. More recently still, advances in natural language processing (NLP) and machine learning techniques have encouraged expectations that manual coding might be dispensed with altogether (Schrodt, 2012).

Yet such a narrow explanation is not entirely satisfying. It ignores both the reasons for the failure of previous endeavours to catch on, insofar as these cannot be reduced to limited computational power alone, and the actual findings of the research programs in question. To give an adequate account of the phenomenon, we must look at whether and how the use of statistical and mathematical techniques and large-N sample sizes has produced new answers to old problems, or opened up areas of inquiry hitherto ignored by more traditional approaches. It is also germane to consider the historical context for the shift under consideration, both internal (to English-language IR scholarship) and geopolitical; finally, at the intersection of these two contextual queries, we can ask how realignments of the post-Cold War period affected the relationship between mainstream international security studies (ISS) and the fissiparous field of peace research, long at the positivist avant-garde (Buzan & Hansen, 2009).

Our research project has proceeded on three fronts. First, we have reconstructed the proximate origins of scientific conflict research and the evolution of the two outlets at the forefront of the field, the *Journal of Conflict Resolution* and the *Journal of Peace Research*, charting the challenges posed to both journals by the intellectual and political shockwaves of the 1960s and 1970s, consequences of the 'failure to predict' the collapse of the USSR for quantifiers in IR (Lebow & Risse-Kappen, 1995), and the subsequent, meteoric ascent of conflict data. Second, we have undertaken the systematic coding and analysis of a corpus of articles drawn from the two journals from 1990 to the present, using a framework that integrates theoretical paradigm(s) referenced, methods and data employed, along with generality of explanation, implied likelihood of future conflict, and normative conclusions. These findings are in the process of being integrated into a searchable ATLAS.ti database, with the ambition of detecting similar patterns in the wider literature. Third, as described in greater detail in the next section, we have constructed a second, thematic corpus – articles on the 'rise of China,' drawn from a sample of quantitative and non-quantitative journals – that has yielded a series of preliminary conclusions from the application of the same analytical framework.

Analysing the 'agency of data' with regards to impact on practitioners' representations of armed conflict and the interpretation of their causes

To analyse the impact of quantitative conflict studies on public and private representations of armed conflict among journalists, NGOs, and political institutions, WP2 relies on a combination of corpus-based quantitative document analysis and qualitative research interviews for the selected case countries: France, Germany, and the UK. In Year 1, WP2 focused on the analysis of the impact of quantitative conflict research on discourses, policy papers, and NGO reports.

The construction of the document corpus

- Media discourses: based on FACTIVA and Europress, corpora for France, Germany, and the UK were completed by selecting periodicals and choosing keywords.
    - French periodicals: *Le Monde* (including lemonde.fr), *Le Figaro*, *Le Parisien / Aujourd'hui en France*, *Le Point*, *L'Express*, *Les Echos* (including lesechos.fr), *La Croix* and *La Croix l'Hebdo*, *Libération* (including liberation.fr), *L'Humanité* and *l'Humanité Dimanche*, *Le Monde diplomatique* (including monde-diplomatique.fr), *Reuters* and *Agence France Presse* (AFP). The selection totals 1084 documents.
    - UK periodicals: *The Times* (including thetimes.co.uk) and *The Sunday Times* (including sundaytimes.co.uk), *The Guardian*, *The Observer*, *The Financial Times* (including ft.com), *The Daily Telegraph*, *The Sunday Telegraph*, *The Economist* (including economist.com), *The Independent*, *New Statesman, The Spectator.* Approximately 2100 documents were selected.
    - German periodicals: *Bild* (including bild.de), *Süddeutsche Zeitung* (including sueddeutsche.de), *Frankfurter Allgemeine Zeitung*, *Handelsblatt* (including handelsblatt.com), *Die Welt* (including welt.de), *taz. die tageszeitung*, *Neues Deutschland* (only the online version: neues-deutschland.de), *Reuters, Deutsche Presse Agentur* (dpa). The selection totals 1197 documents.
    - Keywords: SIPRI or Stockholm International Peace Research Institute or Institut International de Recherche sur la Paix de Stockholm or Correlate of War or ACLED or Armed Conflict Location Event Data Projector or Fragile States Index or Global Peace Index or Conflict Barometer or Uppsala Conflict Data Program or Humanitarian Data Exchange or Index for Risk Management or Eurostats or Global Terrorism Database or Open Situation Room Exchange or Dataminr or Aggle or Ushahidi or HIIK or Heidelberg Institute for International Conflict Research or ICB or International Crisis Behavior or WEIS or World Event Interaction Survey or ICEWS or Integrated Conflict Early Warning System or Global Database for Events Language Tone or GDELT or IISS or International Institute for Strategic Studies or IIES or Institut International d'Etudes Stratégiques.
- Policy papers: **France.** Based on online documents provided by the Ministry of Europe and Foreign Affairs and the Ministry of Armed Forces, 64 documents have been identified that use quantitative data – but not from our pre-established databases – and call for strengthening the use of quantitative data in conflict prevention. From the Ministry of Foreign Affairs: Sector strategy documents (summary documents on France's strategy in various areas of international cooperation); evaluation documents (assessment of sectoral cooperation policies or cooperation projects conducted by the Ministry, carried out at the request of the DGM services, by external consultants); reports and studies; and *fiches* (two-sided sheets presenting the external action carried out by France in a specific area or in collaboration with an organization). For the French Ministry of Armed Forces: observatory reports; prospective and strategic studies. **UK.** Based on online documents from the Ministry of Defence and the Foreign and Commonwealth office, a list of 89 documents using quantitative data has been compiled. Each document highlights the importance of quantitative data, especially regarding conflict analysis, and many mobilise quantitative data from the UN or the OECD. However, only 4 documents (from the Development, Concepts, and Doctrine Centre) use data from conflict-related databases.

Germany. Based on online documents provided by the Federal Foreign Office, the Federal Ministry of Defence, and the Berlin Center for International Peace Operations, 30 documents mentioning at least one of the databases or the use of quantitative data in general were collected. The Centre for International Peace Operations (ZIF) was established in 2002 by the Federal Government and the Bundestag. Its aim is to strengthen Germany's international civilian capacities in crisis prevention, conflict resolution and peacebuilding. The ZIF offers services and expertise (consulting and analysis) in Germany and abroad on the subject of peace operations. The centre manages the recruitment, training and secondment of personnel. For these reasons, ZIF publications have been included here: they represent the majority of the corpus (26 documents). In addition to internal quantitative data produced by the ZIF and other German institutions (such as the GIZ), the two databases that appear to be used are the Global Peace Index and Ushahidi. The documents also show various data from the EU, NATO or the UN. A study of the quantitative data produced by the OECD in the framework of the International Network on Conflict and Fragility (INCAF) is also mentioned. The use of quantitative data by federal institutions is confirmed by the PREVIEW project, although the sources of the data studied are not specified.

Due to the lack of sufficient sources, it appears that lexicometric exploitation of the corpus is of limited interest. Preliminary interviews (carried out by WP III) revealed the use of conflict-related databases in internal ministerial communications, which are classified and do not appear in public documents. The team agreed to use this corpus to prepare for the interviews scheduled for the second year.

- **NGO reports:** annual reports from 17 peacekeeping NGOs were analysed (203 documents), which reveal considerable use of quantitative data but which either did not cite their sources or else cited sources absent from our pre-established list – most of the time, these sources consisted in UN databases or internal data.
  - *List of NGOs:* Act Alliance EU, ACTED, Action contre la faim, ADRA, Care International, La chaine de l'espoir, CICR, Conciliation resources, Handicap International, International Alert, International Crisis Group, Médecins du monde, Médecins sans frontière, Oxfam, Première urgence internationale, Solidarités international, Voice EU.

Collecting practitioner input and disseminating project results and policy advice

The overall objective of the research programme is to develop practical guidance for practitioners involved in conflict management, including diplomats, military officers, journalists, and NGO staff, so as to encourage greater awareness regarding the use of quantitative conflict data and new theoretical insights, as well as vigilance with respect to potential data-engendered biases.

Our intent is to help practitioners in their day-to-day work by fostering a more reflective use of quantitative conflict data in analysis, decision-making, and operations. Taking into account the feedback we gathered in three non-public workshops with NGO workers, military officials, and journalists, WP3 will prepare the major dissemination outcome of DATAWAR, the **Conflict Database Compass** (CDC). The CDC website will be the first publicly accessible and interactive guide to available sources of quantitative conflict data, organized according to specific practitioner needs. Based on a detailed mapping exercise of open-source databases, the tool will gather meta-data for each database and make them accessible through interactive

queries. The CDC will thus assist practitioners, who are often less familiar with the pitfalls of gathering social science data while working under time pressure, in quickly identifying the conflict database that is most reliable and germane to their particular needs. To launch the CDC website and initiate exchange between practitioners and conflict researchers, WP3 will organize a public workshop involving practitioners, academics, and graduate students (Fall 2022).

The final objective of the WP3 work plan is to ensure a wide dissemination of results, training, and advice for civilian and military practitioners. By liaising with individual practitioners (including military planners, NGO staff, and journalists), WP3 aims at stimulating discussion and awareness while helping to disseminate the academic results of the DATAWAR project among these audiences. This will be implemented through two complementary actions: 1) WP3 will organize **a two-day dissemination event**, including a one-day non-public workshop allowing for exchange with stakeholders from the external advisory board as well senior practitioners of conflict management, and a one-day public dissemination conference. 2) In 2023, WP3 will prepare and realize a **two-month training course (8 evening sessions) for practitioners** that should enable participants to experiment with and critically reflect upon different ways of mobilizing quantitative conflict data. This course will be offered at Sciences Po Lille and specifically address military officers, members of civilian crisis management cells, journalists, and humanitarians. The course will be promoted through existing partnerships as well. Additionally, **a specific version** of this course will be **integrated into the MA programme** of Sciences Po Lille.

Preliminary Findings

Is there a deterministic bias among quantitative scholars?

Based on a comparison of the 'causes' of armed conflict in the fifty most-cited articles published in two leading quantitative journals (the *Journal of Conflict Resolution* and the *Journal of Peace Research*) and two more qualitatively oriented journals (*Security Dialogue*, *Millennium*), we have examined the relationship between methodological orientation and determinist visions of war. Overall, we find strong evidence that quantitative articles chiefly mobilize 'material' variables in their research design (such as balance of power, regime type, economic interdependency, and natural resources), while 'interpretative' approaches accord greater consideration to the meaning actors assign to their actions (the construction of identity, norms, and emotions). At the same time, quantitative studies that measure conflictual interactions typically present a more pessimistic vision of conflict-avoidance than sociologically and historically oriented qualitative work. This has led us to the hypothesis that the positivist attempt to establish regularities may in some circumstances favour inflated threat perceptions, both by homogenizing actors and through explicit or implicit meteorological analogies, likening war to natural calamity.

To examine this question concretely, we looked at the specific case of expectations concerning a future military conflagration involving China. We tested the hypothesis that quantitative models based on large case comparisons tend to be more pessimistic and more deterministic in their evaluation of the rise of China, as expressed in the so-called 'Thucydides Trap' (Allison, 2015), the vocabulary of which has increasingly been adopted by US policy-makers as well as IR scholars. Our tentative conclusion is that differences in how the Chinese 'threat' is evaluated and the policy recommendations that ensue reflect not only paradigmatic choices but also more or less positivistic outlooks.

Our corpus included a sample of the most-cited articles in five leading journals referring to the thematic of China and possible conflict with its neighbours, the US, and other countries. The selection principle combined two criteria: journals' ranking according to the Thomson Reuters citation index, and their contrasting epistemological and methodological orientations. The *Journal of Conflict Resolution* and the *Journal of Peace Research* are both widely cited and share an outspoken positivist epistemology, while *Millennium* and *Security Dialogue* are well-established forums for critical, post-positivist scholarship. Finally, we included the *Chinese Journal of International Relations*, an epistemologically catholic journal devoted to the thematic area of interest.

To analyse this corpus, we applied a standardized framework. First, we looked to the production of data: are data 'standardized' via quantification or other casing procedures? Do authors employ data to establish 'regularities' irrespective of temporal and spatial context, or are they sensitive to contextual factors? We assume that authors who standardize data through quantification also tend project past patterns into the future, issuing in a more 'fatalistic' appreciation of threats than approaches that emphasize changes and breaks in the history of IR. Second, we identify the methods used. Do scholars rely on statistical models to verify causal relations? What variables are adduced and how are they coded? Is there a preference for 'measurable' over less measurable factors? Third, we attempt to identify the underlying social ontology of the research programs under investigation. Do positivists, as we expect, privilege structural determinations over actors' agency? How do they qualify social relations: are they reduced to considerations of utility maximization or do they integrate 'social' attitudes and affects? Do they postulate the existence of trends or even laws, including teleological conceptions of progress or decline?

The fourth element of the framework poses two final, decisive questions. First, does the author expect that war or commercial conflict between China and other powers is more or less likely? Second, what kind of policy recommendations, if any, does the author propose? Interrogation on these lines has permitted us to examine the links between methodological positivism, worldview, and framing of international conflict.

Do media and institutions actually use academic conflict data in their output?

Media discourses: The quantitative analysis of the French corpus revealed clear disparities in the use of databases by the media. Figure 1 shows the occurrence frequency of databases in the selected press articles. SIPRI, ACLED and IISS are the three main databases cited, SIPRI representing almost 65% of all occurrences. ICB and FSI are virtually never used.

Distribution of codes (Frequency)
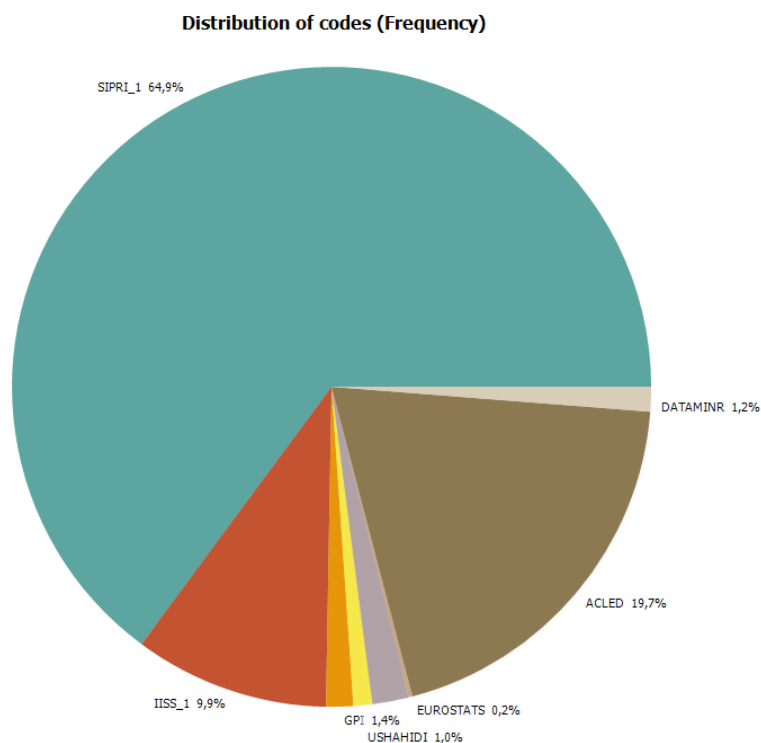
Figure 1: Databases' occurrence frequency in the French corpus.

Media do not seem to use databases in a homogeneous way. Figure 2 shows that *Le Monde* is by far the media referring the most frequently to databases, followed by *Le Figaro* and *Les Echos*. This can partly be explained by the fact that these media are daily publications. It is therefore important to stress that publication frequency does have an impact on the overrepresentation of daily newspapers in this corpus. However, since most media have now developed online outlets (which were computed in this study under the same label as the print versions) publication frequencies do not follow as clear a rhythm as their type initially suggests. The bias implied by the difference in frequency publication is thus partially reduced. Media such as *Le Point*, *Le Parisien* or *L'Express* seem to barely ever use databases on conflict. What is also striking is the relatively small number of occurrences in the corpus spanning over more than thirty years.
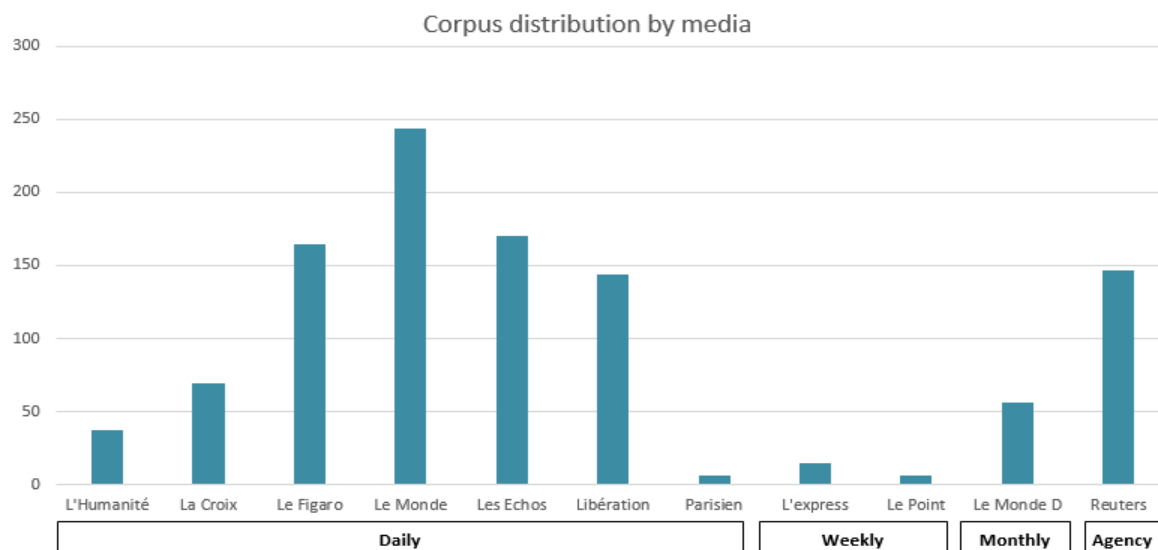
Figure 2: Frequency distribution of media from the French corpus

Looking closer at the distribution by media (Figure 3), it appears that whereas SIPRI is referenced in all media sources, ACLED is mainly mobilised by *Le Monde* and *Libération* while IISS appears more frequently in *Libération*, *Le Figaro* and *Les Echos*.
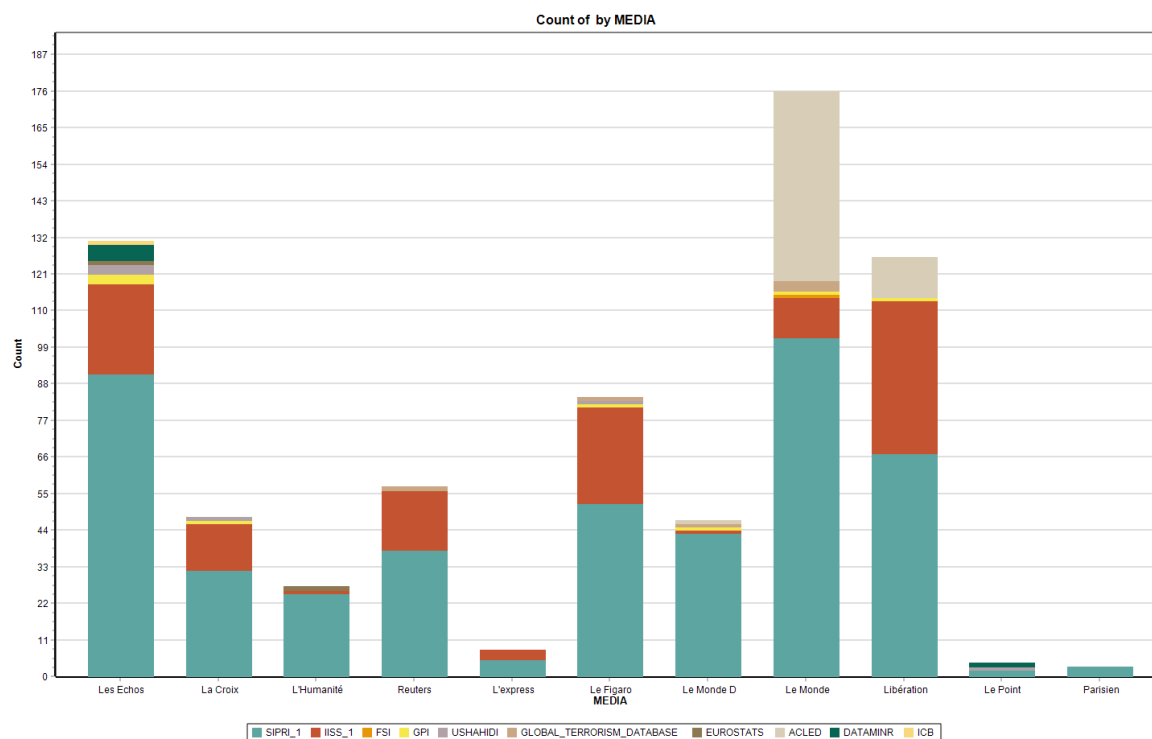

Figure 3: Distribution of databases by media in the French corpus

The general use of databases has grown steadily over the years, reaching its highest point in 2019. However, Figure 4 shows marked increases following 2001 and then following 2016. This seems to parallel the transformations of conflictuality linked to the attacks on the World Trade Center and Pentagon in September 2001 as well as the beginning of the war in Syria and mounting instability in the Middle East.
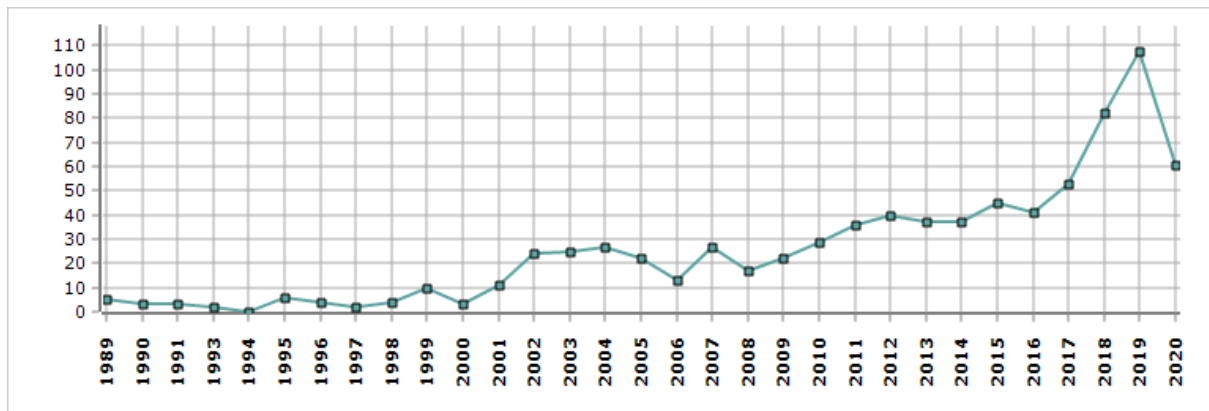

Figure 4: Evolution of database use

Finally, a close look at how the databases are mobilised reveals a focus on the state of military spending rather than the interpretation of armed conflicts. This seems consistent with the overrepresentation of SIPRI, which provides extensive analysis and data related to states' military expenditures.

The descriptive results outlined above raise multiple questions regarding the causal mechanisms underlying disparities in the use of databases. These call for further investigation of the relational affiliations between the newsrooms and the institutes selected for the study, as well as the weight of the different periodicals' political and thematic orientation. Recent developments in data journalism will also be considered.

Preliminary quantitative analysis of the British corpus also indicates that SIPRI and IISS are the most frequently cited databases.

Quantitative analysis of the German corpus leads to similar conclusions as those drawn from the French corpus. First, SIPRI and IISS are the two most-cited databases in a corpus of 1197 articles: 594 mentions for SIPRI, 209 for IISS. The Global Peace Index is the third most-cited database with 26 mentions. Secondly, these references concern global defence spending rather than the interpretation of armed conflict.

The next step will be to continue pursuing quantitative analysis in three directions: distribution of references by publication; type of assessment by geographical area; and nature of reference (explanatory versus prospective use of data, degree of alarmism in threat assessments and forecasts issuing from the interpretation of data).

Beyond the 'data pessimists' – 'data optimists' divide: The case of practitioners

Scholarship on the use of quantitative conflict data by practitioners is structured by a divide that applies to a greater or lesser degree across the different actor categories analysed by DATAWAR. This has already been observed for the area of international peacekeeping by Roger Mac Ginty, who notes that 'significant

debate exists between what might be called digital optimists and digital pessimists' (Mac Ginty, 2017, p. 698), as well as for the area of conflict analysis by Larissa Fast, who emphasizes the existence of an opposition between the 'perspectives of the "data enthusiasts" and the concerns of the "data sceptics"' (Fast, 2017, p. 709).

In our preliminary practitioner workshops and informal encounters, conducted to date (as of April 2021) with practitioners from humanitarian organizations, we have been able to confirm the existence of this bipartite opposition – with a few nuances, however. First, the 'pessimist-optimist' can be found even within organizations, reflecting a struggle between competing professional cultures rather than top-down organisational decisions. Second, even among actors that clearly qualify as 'data optimists,' insofar as they highlight the potential of quantitative data to improve decision-making, early hopes in the forecasting capability of quantitative analysis seem to have been tempered. The 'optimists' with whom we spoke generally emphasize that quantitative data should be seen as a useful complement to existing, qualitative assessment procedures, not a satisfactory replacement for the latter. In particular, it appears that the expectations placed on predictive and early-warning approaches have been disappointed because it is simply too difficult to identify the causal mechanisms that would enable reliable data-based forecasting. Third, we were surprised that most organisations rely on quantitative data collected as part of their own activities and do not make use of academic conflict databases, which are perceived as insufficiently precise and up-to-date. From our early consultations as well as the media analysis conducted by WP2, it appears that ACLED is the only 'classical' database to be used across different practitioner groups. Furthermore, although some organizations endeavour to promote exchange between scholars and practitioners, such exchanges tend to be ineffective due to the divergent timelines and priorities of researchers and practitioners.

Ongoing consultations with practitioners indicate a growing interest as well as genuine concern over the challenges posed by forms of data-dependency in multiple professional sectors. Information dominance, a contemporary conceptual development made possible by increasing possibilities for real-time management of data flows, is at the heart of an ongoing transformation of military institutions. For their civilian counterparts, the risk of being instrumentalised in decision-makers' quest to resolve complex contingency needs underlines the threat as well as the promise of big data as a means of exercising operational power.

Perspectives

At this stage, we can identify three major initiatives in the short term. First, the articulation between the three work-packages will be reinforced. For instance, reference to historical analogies to make sense of contemporary data is a transversal concern for the perception – whether alarmist or irenic – of international conflict by scholars, soldiers, diplomats, NGO analysts, and journalists. Second, in order to decode varying attitudes among different communities of users it is necessary to understand how database functionality is integrated within a broad range of outlooks, personal and professional orientations (styles of argumentation and justification, worldview…). Third, interactions between the team and practitioners have underscored a strong interest in conflict databases, especially on the part of military officers. Changes in the global system affect the demand for knowledge about war; renewed interest in

high-intensity interstate warfare, for example, undoubtedly has significance for the production and use of data. What part exactly does the transformation of strategic relationships play? This aspect was not directly considered in our initial project, whereas it frequently appears in the exploratory interviews conducted so far; additional reflection on the subject is in order.

References

Allison, G. (2015). The Thucydides Trap: Are the US and China Headed for War? *The Atlantic*, (24 September 2015). Available online.

Baele, S. J., Balzacq, T., & Bourbeau, P. (2018). Numbers in global security governance. *European Journal of International Security*, 3(1), 22-44.

Baillat, A., Emprin, F., & Ramel, F. (2016). Des mots et des discours : du quantitatif au qualitatif. In G. Devin (Ed.), *Méthodes de recherche en relations internationales* (pp. 227-246). Paris: Les Presses de SciencesPo.

Balzacq, T. (2014). The significance of triangulation to critical security studies. *Critical Studies on Security*, *2*(3), 377-381.

Bauer, H., & Brighi, E. (Eds.). (2008). *Pragmatism in International Relations*. London: Routledge.

Beerli, M. J. (2017). Legitimating Organizational Change through the Power of Quantification: Intra-Organizational Struggles and Data Deviations. *International Peacekeeping, 24*(5), 780-802.

Buzan, B., & Hansen, L. (2009). *The Evolution of International Security Studies*. Cambridge: Cambridge University Press.

Bruno, I., Didier, E., & Prévieux, J. (2014). *Statactivisme : comment lutter avec des nombres*. Paris: Zones.

Clément, M., & Sangar, E. (Eds.). (2018). *Researching Emotions in International Relations: Methodological Perspectives on the Emotional Turn*. Cham: Palgrave Macmillan.

Colonomos, A. (2016). *Selling the future: The perils of predicting global politics*. Oxford: Oxford University Press.

De Franco, C., & Meyer, C. O. (Eds.). (2011). *Forecasting, warning, and responding to transnational risks*. Houndmills, Basingstoke: Palgrave Macmillan.

Der Derian, J. (2013). From War 2.0 to quantum war: The superpositionality of global violence. *Australian Journal of International Affairs, 67*(5), 570-585.

Desch, M. C. (2019). *Cult of the Irrelevant: The Waning Influence of Social Science on National Security*. Princeton: Princeton University Press.

Dieckhoff, M., Martin, B., & Tenenbaum, C. (2016). Classer, ordonner, quantifier. In G. Devin (Ed.), *Méthodes de recherche en relations internationales* (pp. 247-266). Paris: Les Presses de SciencesPo.

Duursma, A., & Read, R. (2017). Modelling Violence as Disease? Exploring the Possibilities of Epidemiological Analysis for Peacekeeping Data in Darfur. *International Peacekeeping, 24*(5), 733-755.

Eberwein, W.-D., & Chojnacki, S. (2001). *Scientific necessity and political utility: A comparison of data on violent conflicts*. WZB Discussion Paper.

Fast, L. (2017). Diverging Data: Exploring the Epistemologies of Data Collection and Use among Those Working on and in Conflict. *International Peacekeeping, 24*(5), 706-732.

Fioramonti, L. (2014). *How numbers rule the world: The use and abuse of statistics in global politics*. London: Zed Books.

Franz, N. (2017). Targeted killing and pattern-of-life analysis: Weaponised media. *Media, Culture & Society, 39*(1), 111-121.

Freedman, L. (2017). *The future of war: A history.* New York: PublicAffairs.

Girod, D. M. (2015). *Explaining postconflict reconstruction*. New York: Oxford University Press.

Gleditsch, K. S., Metternich, N. W., & Ruggeri, A. (2014). Data and progress in peace and conflict research. *Journal of Peace Research, 51*(2), 301-314.

Greenhill, K. M. (2010). Counting the cost: The politics of numbers in armed conflict. In P. Andreas & K. M. Greenhill (Eds.), *Sex, drugs and body counts: The politics of numbers in global conflict and crime* (pp. 127–158). New York: Cornell University Press.

Hansen, H. K., & Porter, T. (2012). What Do Numbers Do in Transnational Governance? *International Political Sociology, 6*(4), 409-426.

Harbom, L., Melander, E., & Wallensteen, P. (2008). Dyadic Dimensions of Armed Conflict, 1946—2007. *Journal of Peace Research, 45*(5), 697-710.

Hartog, F. (2003). *Régimes d'historicité: Présentisme et expériences du temps*. Paris: Editions du Seuil.

Himelfarb, S. (2014). Can Big Data Stop Wars Before They Happen? *Foreign Policy*, (25 April 2014). [Available online](#).

Kitchin, R. (2014). *The data revolution: big data, open data, data infrastructures & their consequences*. Los Angeles: SAGE Publications.

Kratochwil, F. (2006). History, Action and Identity: Revisiting the 'Second' Great Debate and Assessing its Importance for Social Theory. *European Journal of International Relations, 12*(1), 5-29.

Latour, B. (1987). *Science in action: how to follow scientists and engineers through society*. Cambridge: Harvard University Press.

Lebow, R.N., & Risse-Kappen. (Eds.). (1995). *International Relations Theory and the End of the Cold War*. New York: Columbia University Press.

Li, Q. (2019). The Second Great Debate Revisited: Exploring the Impact of the Quantitative-Qualitative Divide in International Relations. *International Studies Review, 21*, 447-476.

Lindemann, T. (2011). *Causes of war: The struggle for recognition*. Colchester: ECPR Press.

Lindemann, T. (2016). La construction de l'objet et la comparaison dans l'études des relations internationales. In G. Devin (Ed.), *Méthodes de recherche en relations internationales* (pp. 39-56). Paris: Les Presses de Sciences Po.

Mac Ginty, R. (2017). Peacekeeping and Data. *International Peacekeeping, 24*(5), 695-705.

Mackenzie, D. (1991). Comment faire une sociologie de la statistique.... In M. Callon & B. Latour (Eds.), *La science telle qu'elle se fait* (pp. 200-261). Paris: La Découverte.

Mearsheimer, J. J., & Walt, S. M. (2013). Leaving Theory Behind: Why Simplistic Hypothesis Testing is Bad for International Relations. *European Journal of International Relations, 19*(3), 427-457.

Meyer, C. O., De Franco, C., & Otto, F. (2019). *Warning about war: conflict, persuasion and foreign policy*. Cambridge / New York: Cambridge University Press.

Müller, T. R., & Bashar, Z. (2017). 'UNAMID Is Just Like Clouds in Summer, They Never Rain': Local Perceptions of Conflict and the Effectiveness of UN Peacekeeping Missions. *International Peacekeeping, 24*(5), 756-779.

Porter, T. (1995). *Trust in Numbers*. Princeton: Princeton University Press.

Schrodt, P. A. (2012). Precedents, Progress, and Prospects in Political Event Data. *International Interactions: Empirical and Theoretical Research in International Relations, 38*(4), 546-569.

Schrodt, P. A. (2014). Seven Deadly Sins of Contemporary Quantitative Political Analysis. *Journal of Peace Research, 51*(2), 287-300.

Supiot, A. (2015). *La gouvernance par les nombres : cours au Collège de France, 2012-2014*. Paris: Fayard.

Ward, M. D., Metternich, N. W., Dorff, C. L., Gallop, M., Hollenbach, F. M., Schultz, A., & Weschle, S. (2013). Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction. *International Studies Review, 15*(4), 473-490.