

A future framework for sharing and displaying spatio-temporal indicators

Mayeul KAUFFMANN (mayeul.kauffmann@ecoledelapaix.org)

This is a draft (v. 0.1.2-en - 15 June 2007). Comments are warmly welcomed.

Une version en français est disponible.

Note: *This project was initially aimed at integrating databases related to armed conflict, but it may be expanded to include human rights and environment areas as well, with few changes.*

This project aims at building an information system for peace research science, targeted both at researchers and at a much broader audience. The information system would be mainly web based, serving as a portal to peace issues, but a standalone, desktop system could be derived from it, for advanced research in peace science and international issues. An important part of the project is to effectively use this tool for a quantitative peace research analysis.

This project would be made of the following components:

1. **Merging:** Merge the available databases that are useful for peace research;
2. **Displaying:** Display these data in an interactive, graphical form (over the web or as a standalone application), with one-click access to related textual documents;
3. **Using it:** Perform a quantitative peace research analysis using this tool (e.g. a spatiotemporal analysis of the causes of civil wars).

Targeted audiences

The information system would be targeted to three different groups:

1. Any person interested in peace and international relations.
2. Researchers in peace research and international relations.
3. Students in these fields.

The wider targeted audience would be any person feeling concerned about peace and international relations. This would include simple citizens, journalists, NGO's activists, politicians, diplomats...

With respect to students in peace research and international relations, the information system could be used as a research and educative device.

Researchers would certainly have more advanced requirements in terms of functionalities.

The information system could have 3 levels of use:

1. **"Discover"**. This may be a short on-line gallery of simple maps and diagrams, preconfigured and immediately visible by the visitor, nearly screenshots. The gallery will be chosen to focus on various important issues in international affairs and would aim at raising the awareness of the visitor. A few lines of comments could help the user interpret the pictures. The interface should have minimal functions, maybe just "zoom" and "pan".

2. “**Standard**”. This should allow most common functionalities: zoom, pan, change date or time interval, choosing layers from the available datasets, etc.

3. “**Expert**”. This would allow retrieving numeric data from the database, doing custom queries, configuring the display of layers (choosing colours and symbols), saving parameters for future display of a given map or diagram, maybe adding one’s own dataset. This would be targeted to researchers. For technical reasons, this will require creating an account (which may be given for free). Alternatively, some of these functions (e.g. adding one’s own dataset, retrieving numeric data) may only be available on the standalone version of the information system.

In order to facilitate access:

The web interface should be usable with most common web browsers (IE6, Firefox, Safari, Opera...).

The web interface as well as the standalone interface should be multilingual.

We now describe the three main components of the project.

1. Merging current peace research databases

Currently, when doing a quantitative research project, peace research specialists can download the needed databases in just a few minutes, but merging them can be a very complex and cumbersome task, requiring expertise not only on the way the data were coded, but also in database management. As a simple and straightforward example of this, the fact that not all datasets use the same sets of countries is in itself a difficulty and may even lead to mistakes (not even to mention the different names for the same countries in different datasets, or the different ways of handling secession or unification of states). The problems become even more complicated when one tries to use data *at the conflict level* (for instance, one conflict in a given dataset may be split in several conflicts in another one).

This project may include some or all of the following databases, as well as any other relevant database.

Tentative list of databases:

- UCDP (Uppsala Conflict Database Project, Uppsala Univ. and PRIO)
- COW (Correlates of War, University of Michigan)
- databases from the University of Maryland (ICB [International Crisis Behaviour], MAR [Minorities at Risk], Polity IVd [daily data on political regimes], etc.)
- GRIP’s data (contribution to world peace, ...)
- KOSIMO (Konflikt-Simulations-Modell, Heidelberg Institute of International Conflict)
- SIPRI’s datasets
- databases on arms transfers (NISAT’s data on small arms, ...)
- datasets on arms expenditures
- UNDP’s socio-economic data (HDI, ...)
- UNHCR’s data on refugees
- a subset of World Bank’s *World Development Indicators (W.D.I.)*
- and more if possible.

Inclusion on this list will depend on agreements with partners (for copyrighted resources), on financial resources (merging datasets is time consuming), on usefulness for researchers and on potential interest of the wider audience.

Substantial subsets of the datasets from the UNDP, UNHCR and World Bank are already

available on the Internet. These organizations might give us the right to use part of their data.

For the case of non-free databases (e.g. the *W.D.I.* World Bank CD-ROM), the information system could include ready-to-use link-tables and pre-configured procedures, which would be useful only to owners of the datasets. This would be targeted to researchers.

Of course, the sources of all materials will be clearly acknowledged, with a direct link to the data producer's website. This could be a way for research centres to make their own work known.

This relational database management system (RDBMS) would consist of 5 components:

- a database engine,
- the existing above-mentioned datasets,
- link-tables to enable merging these datasets,
- procedures (SQL queries) to import, organize, merge and query these datasets,
- a G.U.I. (graphical user interface).

Ideally, this should be available both as a standalone application (to be run on a single desktop) and on a web server (for data to be managed and retrieved through a web browser).

2. Display spatiotemporal and textual data

The goal is to be able to easily use and display any data of the above mentioned datasets, and link them with textual data.

This part will be made of three sub-components, the first favouring the spatial component (A), the second focusing on the temporal component (B), and the third linking them to textual data (C).

A. Spatial representations

The information system will display data in the form of on-line maps. These maps will be interactive, and will support zooming, panning, selecting data (layers) to display, etc.

These maps may show various contents:

- political regimes shown as different colours
- geographical extent of civil war at a given date
- origin and destination of arms transfers
- level of military expenditures
- socio-economic data (HDI, GDP..)
- etc.

For spatial data with a temporal component, a slider should let the user change the year (or date) of the map. Two sliders may be used for displaying short-length events that occurred during a given time interval, one slider for the start date of the time interval displayed, and the second for the end date.

B. Time-line diagrams

Maps alone cannot always efficiently show a succession of events, especially over a long period. At best, the user sees a succession of maps (thanks to sliders), but cannot see the whole picture at a glance. Maps should be supplemented by time-line diagrams (aka chronologies) whenever it is appropriate.

Data on these diagrams could be displayed using:

- symbols (for punctual events)
- horizontal bars (for time intervals)
- horizontal *coloured* bars (the colour being related to the value of discrete variables)
- line charts.

Selected on-line maps and time-line diagrams that were made for this preliminary study are given in the accompanying document (Annex 1).

C. Textual data

“A picture is worth a thousand words”, but often a thousand words are not enough when you want more in-depth knowledge of some events. Several of the above-mentioned **databases include textual description of events** (for instance, a few lines briefly describing a conflict). At least, such data should be made available from the diagrams, for instance by clicking on the symbol which stands for the start of a conflict in a time-line diagram, or on a conflict area in a map. Very synthetic information may also be given with mouse roll-over (*i.e.* without clicking).

A more advanced tool may also include a link to on-line data. Two kinds of on-line data could be useful:

First, **bibliographical databases**: clicking on an event may display the references of books, theses and articles written on this event. These data could be retrieved using standard library internet protocols, enabling the display of summaries. Depending on user preferences or location, the system might display only references from the local library, or even from the user’s personal database.

The second type of data could be existing on-line **web pages**. This component is really an optional one, since the accuracy of web pages is sometimes low. A possible use would be to link the events with the events recorded in **Wikipedia’s chronologies**. (Wikipedia is not an authoritative source and should sometimes be used with caution; however, the review *Nature* [Vol 438, 15 December 2005] asked experts to compare a selection of articles from it and from the *Encyclopaedia Britannica*, and found that the accuracy of Wikipedia was nearly the same as that of *Britannica*.)

All these components would constitute an information system for peace research.

3. Using it: Doing a quantitative peace research analysis using this tool

The objective of this part will be manifold:

- to test the information system for research purposes and improve it,
- to show the usefulness of the information system,
- to do a research work which is worthy in itself.

A proposed research program could be a quantitative study on civil wars. The main focus would be the causes of civil wars and the factors that influence their duration and their diffusion. This would require to closely integrate the temporal and spatial components of the study. One of the objective of the study could be to assess so-called “structural strategies” (for instance focusing on “human security” in order to prevent armed conflicts), using alternative indicators such as socio-economic data from the UNDP. It may also contribute to the knowledge of the geographic analysis of civil wars and help improve existing methodologies.

Concerning the analytical research framework, it should be aimed at merging some current geographical and temporal statistical methods.

As an example, some methods available for the quantitative study of civil war are discussed in the following two papers:

- Buhaug, Halvard and Päivi Lujala (2005), "Accounting for Scale: Measuring Geography in Quantitative Studies of Civil War", *Political Geography* 24(4): 399-418.
- Kauffmann, Mayeul. (2005), "Short Term and Event Interdependence Matter: A Political Economy Continuous Model of Civil War under IFI programs", communication at the Third ECPR General Conference, Budapest, September. (A draft was presented at the Jan Tinbergen Peace Science Conference in June 2005).

In addition, the methodology may include other spatiotemporal statistical methods (such as Gibbs Point Processes and similar methods). See for instance the works carried out at the Labsad laboratory in Grenoble:

Bertin, Etienne, Jean-Michel Billiot and Rémy Drouilhet (2004). "Phase-Transition in the Nearest-Neighbour Continuum Potts Model", *Journal of Statistical Physics*, 114, 79-100.

Many tasks required for part 1 (merging datasets) and 2 (displaying data) are required before such an analysis. Part 1 and 2 may even be seen as an extension and a diffusion of works made to prepare part 3. Instead of carrying this research (part 3) using *ad hoc* solutions for problems related to part 1 and 2, this entire project can be seen as a generalization and rationalization of the preliminary work that is mandatory for such a quantitative peace research study. Doing this preliminary work in a way that other researchers (and even a broader audience) could use it is likely to foster transparency in this research process. This may increase its quality and credibility, and encourage debate.

Preliminary steps

The first four steps would be to constitute a team, to establish partnerships, to define precise project specifications and to choose a development framework. For the major part of these steps, they would be probably carried out in this order. However, there should be significant time overlaps between each step and the next one.

Constitute a team

This project will require a close cooperation between various people:

- a project leader,
- an international scientific committee,
- researchers (especially for part 3),
- a specialist in peace research databases,
- contact persons (in particular, maintainers of the various datasets),
- a programmer,
- a specialist in software license and copyright issues,
- beta testers (testing for ergonomics, for good functioning on various platforms, browsers and places...),
- the peace research community.

Some of the members of the team may have several functions, depending on available resources and competence.

Building partnerships

In addition to the partnerships implied by the constitution of the project team, specific partnerships should be built with respect to:

- funding.
- data providers (universities and research centres, international institutions, NGOs...). The question of copyrighted materials will be crucial. In addition, whenever possible, these partnerships will include the help of a contact person for each data source, to help resolve problems when merging the data (*e.g.* in case of ambiguous data documentation).
- web hosting (*e.g.* at a university or a research centre).

Define precise project specifications

This step implies defining a core set of functionalities (part 1 and 2) and the precise scope of the research program (part 3), as well as a tentative milestone date.

For parts 1 and 2, this step will require to define the core sets of functionalities wanted for version 1 of the information system (including the list of merged datasets). These will depend on:

- functionalities and datasets required for part 3,
- wish and advice expressed by members of the peace research community (a questionnaire may be circulated),
- technical ease and feasibility,
- agreements with data copyright owners,
- funding.

The precise scope of the research program (part 3) should be defined by the researchers in agreement with the scientific committee and the project leader. The datasets and functionalities required for part 3 should be included in parts 1 and 2.

As a very tentative milestone, version 1 of the information system may be wanted **6 to 10 months** after the start of the project (depending on the specifications and available resources). It may consist of all or a subset of the above mentioned databases, and of some of the functionalities (“Discover” level + much of “Standard” level if possible).

Choose a development framework

Defining the development framework (*i.e.* the programming language(s), software, IDE, libraries...) of the information system will require to carry on an in-depth comparative study of available technical alternatives. A preliminary study on this has already started; it shows that there are promising alternative development frameworks, several of them being completely based on open source/free software. A preview of what they can offer is shown in the attached document (Annex 1). It is suggested here that the information system should only use free *and* open source software, and maybe released under an open source license, *e.g.* GNU GPL or CeCILL (this does *not* presuppose releasing the datasets themselves under a free license such as the GFDL).