

Peace PRISM

Peace Research and International Statistics Mapper

Mayeul KAUFFMANN (mayeul.kauffmann@upmf-grenoble.fr)

Ceci est un document de travail (v. 0.1.1-fr – 17 juin 2006).

Les commentaires sont bienvenus.

This draft is available in English, too.

Le projet *Peace PRISM* vise à construire un système d'information pour les recherches en sciences de la paix (*peace research*), système destiné aussi bien aux chercheurs qu'à un public beaucoup plus large. Ce système d'information serait principalement accessible par internet, servant de portail à des questions relatives à la paix ; mais un système autonome (pour ordinateur de bureau, sans accès à internet) pourrait en être dérivé, pour des recherches avancées en sciences de la paix et en questions internationales. Une part importante du projet est d'utiliser effectivement cet outil pour une analyse quantitative en sciences de la paix.

Ce projet serait constitué des éléments suivants :

1. **Fusion** : Fusionner les bases de données existantes utiles en sciences de la paix ;
2. **Affichage** : Afficher ces données sous forme graphique et interactive (sur internet ou de manière autonome), avec un accès d'un seul clic à des informations textuelles ;
3. **Utilisation** : Utiliser cet outil pour réaliser une étude quantitative en sciences de la paix (par exemple une analyse spatio-temporelle des causes des guerres civiles)

Public visé

Le système d'information viserait trois groupes différents

1. Toute personne intéressée par la paix et les relations internationales
2. Les chercheurs en sciences de la paix et en relations internationales
3. Les étudiants de ces disciplines.

Le public le plus large serait toute personne se sentant concernée par la paix et les relations internationales. Cela inclurait de simples citoyens, des journalistes, des militants d'ONG, des hommes politiques, des diplomates...

Concernant les étudiants en sciences de la paix et en relations internationales, le système d'information pourrait être utilisé comme un outil éducatif et de recherche.

Les chercheurs auront certainement des exigences plus avancées en termes de fonctionnalités.

Le système d'information pourrait avoir 3 niveaux d'utilisation :

1. « **Découverte** ». Il pourrait s'agir d'une courte galerie internet de cartes et de diagrammes simples, préconfigurés et immédiatement visibles par le visiteur, presque des copies d'écran. La galerie mettra en exergue quelques questions importantes en relations internationales et visera à sensibiliser le visiteur. Quelques lignes de commentaires pourraient l'aider à interpréter les images. L'interface devrait avoir des fonctionnalités minimales, peut-être seulement « zoom » et « panoramique » (déplacement).
2. « **Standard** ». Ce niveau donnerait accès aux fonctionnalités les plus courantes : zoom, panoramique, changer la date ou l'intervalle de temps, choisir les couches parmi les données disponibles, etc.
3. « **Expert** ». Ce niveau permettrait de récupérer des données numériques de la base de données, de faire des requêtes personnalisées, de configurer l'affichage des couches (choisir les couleurs et

symboles), d'enregistrer les paramètres pour l'affichage ultérieur d'une carte ou d'un diagramme, peut-être d'ajouter sa propre série de données. Cet outil serait destiné aux chercheurs. Pour des raisons techniques, cela nécessitera de créer un compte personnel (éventuellement à titre gratuit). Ou bien, certaines de ces fonctions (par exemple ajouter ses propres données, récupérer des données numériques) pourraient n'être disponibles que sur la version autonome du système d'information.

Afin de faciliter l'accès au système :

L'interface internet devrait être utilisable avec les navigateurs les plus fréquents (Internet Explorer 6, Firefox, Safari, Opera...).

L'interface internet de même que l'interface autonome devrait être multilingue.

Nous décrivons maintenant les trois principaux aspects du projet.

1. Fusionner les bases de données existantes en sciences de la paix

Actuellement, lors d'un projet de recherches quantitatives en sciences de la paix, les chercheurs peuvent télécharger les bases de données complémentaires en quelques minutes à peine, mais les fusionner peut être une tâche extrêmement complexe et fastidieuse, nécessitant une expertise non seulement sur la façon dont les données ont été codées, mais aussi dans la gestion des bases de données. Pour prendre un exemple simple, le fait que les différentes séries de données n'utilisent pas le même ensemble de pays est en soit une difficulté et peut même conduire à des erreurs (sans parler de la multiplicité de noms d'un même pays dans les différentes bases, ou les diverses façons de gérer la sécession ou l'unification des États). Le problème devient encore plus complexe lorsqu'on essaie d'utiliser les données *sur les conflits* (par exemple, un conflit dans une base peut être découpé en plusieurs conflits dans une autre).

Ce projet pourrait inclure les bases de données suivantes (ou une partie d'entre-elles), ainsi que toute autre base pertinente.

Liste préliminaire de bases de données

- UCDP (*Uppsala Conflict Database Project*, Uppsala Univ. et PRIO)
- COW (*Correlates of War*, University of Michigan)
- bases de l'University of Maryland (ICB [*International Crisis Behaviour*], MAR [*Minorities at Risk*], Polity IVd [données journalières sur les régimes politiques], etc.)
- données du GRIP (contribution à la paix mondiale, ...)
- KOSIMO (Konflikt-Simulations-Modell, *Heidelberg Institute of International Conflict*)
- données du SIPRI
- bases sur les transferts d'armes (base NISAT sur les armes légères, ...)
- bases sur les dépenses militaires
- données socio-économiques du PNUD (IDH, ...)
- données du HCR sur les réfugiés
- une partie des *World Development Indicators (W.D.I.)* de la Banque mondiale
- et plus si possible.

L'inclusion dans cette liste dépendra des accords avec les partenaires (pour les données sous copyright), des ressources financières (fusionner des données est coûteux en temps), de l'utilité pour les chercheurs et de l'intérêt potentiel du grand public.

Une partie non-négligeable des données du PNUD, du HCR et de la Banque mondiale est déjà disponible sur internet. Ces organisations nous accorderont peut-être l'autorisation d'utiliser une partie de leurs données.

Pour le cas des bases payantes (par exemple le CD-ROM *W.D.I.* de la Banque mondiale), le système d'information pourrait inclure des tables-pivots et des procédures prêtes à l'emploi, qui ne seraient utiles qu'aux possesseurs des données. Cette fonctionnalité serait destinée aux chercheurs.

Bien sûr, les sources de toutes les données seront clairement mentionnées, avec un lien direct vers le site internet de l'organisme qui en est l'auteur. Cela pourrait être un moyen pour des centres de

recherche de faire connaître leurs travaux.

Ce système de gestion de bases de données relationnelles (SGBDR) sera constitué de 5 éléments :

- un moteur de base de données,
- les bases de données existantes (mentionnées ci-dessus),
- des tables-pivots permettant de fusionner ces données,
- des procédures (requêtes SQL) pour importer, organiser, fusionner et interroger ces données,
- une interface graphique.

Idéalement, ceci devrait être disponible aussi bien en tant qu'application autonome (exécutable sur un ordinateur de bureau) qu'à partir d'un serveur internet.

2. Afficher des données spatio-temporelles et textuelles

L'objectif est d'être capable d'utiliser et d'afficher facilement les données mentionnées ci-avant, et de les relier à des informations textuelles.

Cette partie sera constituée de trois sous-composants, le premier favorisant la dimension spatiale (A), le second se concentrant sur la dimension temporelle (B), et le troisième reliant ces dimensions à des informations textuelles (C).

A. Représentations spatiales

Le système d'information affichera les données sous la forme de cartes en ligne. Ces cartes seront interactives, et permettront le zoom, le panoramique, la sélection des données (couches) à afficher, etc.

Ces cartes pourront montrer divers contenus :

- les régimes politiques représentés par différentes couleurs
- l'extension géographique des guerres civiles à une date donnée
- l'origine et la destination des transferts d'armes
- le niveau des dépenses militaires
- des données socio-économiques (IDH, PIB...)
- etc.

Pour les données spatiales ayant aussi une dimension temporelle, un curseur devrait permettre à l'utilisateur de changer l'année (ou la date) de la carte. Deux curseurs pourraient être utilisés pour afficher les événements (de courte durée) ayant eu lieu durant un intervalle de temps donné (un curseur pour le début, un pour la fin).

B. Diagrammes temporels

Des cartes ne sont pas toujours, à elles seules, suffisantes pour montrer une succession d'événements, en particulier sur une longue période. Au mieux, l'utilisateur voit une succession de cartes (grâce à un curseur), mais il ne peut pas avoir de vision d'ensemble d'un seul coup d'oeil. Les cartes devraient être complétées par des diagrammes temporels (des chronologies) quand cela est approprié.

Les données sur ces diagrammes pourraient être affichées par :

- des symboles (pour les événements ponctuels),
- des barres horizontales (pour les intervalles de temps),
- des barres horizontales *colorées* (la couleur dépendant de la valeur de variables discrètes),
- des graphiques (courbes).

Une sélection des diagrammes temporels et des cartes en ligne qui ont été réalisés pour cette étude préliminaire est présentée dans le document d'accompagnement (Annexe 1).

C. Données textuelles

« Une image vaut mille mots », mais souvent mille mots ne suffisent pas pour une connaissance approfondie de certains événements. Plusieurs **bases de données** mentionnées ci-dessus **incluent**

une description textuelle des évènements (par exemple quelques lignes décrivant brièvement un conflit). Au minimum, ces informations devraient être accessibles à partir des diagrammes, par exemple en cliquant sur le symbole qui représente le début d'un conflit dans un diagramme temporel, ou sur une zone de conflit sur une carte. Des informations très synthétiques pourraient aussi être données par simple survol de la souris (c'est-à-dire sans cliquer).

Un outil plus avancé pourrait inclure un lien vers des données en ligne. Deux types de données pourraient être utiles :

D'abord, des **bases de données bibliographiques** : cliquer sur un événement permettrait d'afficher les références de livres, thèses et articles traitant de cet événement. Ces données pourraient être récupérées en utilisant les protocoles internet standards des bibliothèques, permettant l'affichage des résumés. En fonction des préférences de l'utilisateur ou de sa localisation, le système pourrait afficher seulement les références de la bibliothèque locale, ou même de la base de données personnelle de l'utilisateur.

Le second type de données pourrait être constitué de **pages internet** déjà en ligne. Cet élément est optionnel dans la mesure où la fiabilité des pages internet est parfois limitée. Une utilisation possible serait de relier les événements de la base avec ceux des **chronologies Wikipédia**. (Wikipédia n'est pas une source qui fait autorité et doit parfois être utilisée avec précaution ; cependant, la revue *Nature* [vol. 438, 15 décembre 2005] a demandé à des experts de comparer une sélection d'articles de Wikipédia et de *l'Encyclopaedia Britannica*, et a conclu que la fiabilité de Wikipédia était presque du même niveau que celle de *Britannica*.)

Tous ces composants constitueront un système d'information pour les recherches en sciences de la paix.

3. Utilisation : faire une recherche quantitative en sciences de la paix en utilisant cet outil

L'objectif de cette partie est multiple :

- tester le système d'information à des fins de recherches et l'améliorer,
- montrer l'utilité du système d'information,
- faire un travail de recherche ayant une valeur intrinsèque, indépendamment des éléments ci-dessus.

Une proposition de programme de recherche est l'analyse quantitative des guerres civiles. Le principal objet d'analyse serait les causes des guerres civiles et les facteurs qui influencent leur durée et leur diffusion. Cela nécessiterait d'intégrer étroitement les dimensions temporelles et spatiales de l'étude. Un des objectifs de l'étude pourrait être d'évaluer les « stratégies structurelles » (avec par exemple une attention particulière pour la « sécurité humaine » dans la prévention des conflits armés), en utilisant des indicateurs alternatifs comme les données socio-économiques du PNUD. Cela pourrait aussi contribuer à l'étude géographique des guerres civiles et à l'amélioration des méthodes existantes.

Concernant le cadre analytique, il devrait viser à fusionner certaines méthodes statistiques spatiales et temporelles actuelles.

À titre d'exemple, certaines méthodes disponibles pour l'analyse quantitative des guerres civiles sont discutées dans les deux documents suivants :

- Buhaug, Halvard et 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 à la *Third ECPR General Conference*, Budapest, septembre. (Une version de travail a été présentée à la *Jan Tinbergen Peace Science Conference* en juin 2005.)

En outre, la méthodologie pourrait inclure d'autres techniques statistiques spatio-temporelles

(comme les processus ponctuels de Gibbs et des méthodes similaires). Cf. par exemple les travaux effectués au laboratoire Labsad de Grenoble :

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

De nombreuses tâches des parties 1 (fusionner) et 2 (afficher) sont nécessaires avant une telle analyse. Les parties 1 et 2 peuvent même être vues comme une extension et une diffusion des travaux effectués pour préparer la partie 3. Au lieu d'effectuer cette recherche (partie 3) en utilisant des solutions *ad hoc* aux problèmes liés aux parties 1 et 2, l'intégralité du projet peut être vue comme une généralisation et une rationalisation du travail préalable indispensable à la réalisation d'une telle étude. Faire ce travail préliminaire de sorte que d'autres chercheurs (et même un public plus large) puissent s'en servir est susceptible de promouvoir la transparence du processus de recherche. Cela pourrait accroître sa qualité et sa crédibilité, et encourager le débat.

Étapes préliminaires

Les quatre premières étapes devraient être de constituer une équipe, d'établir des partenariats, de définir précisément les spécifications du projet et de choisir un cadre de développement de l'application. Pour l'essentiel, ces étapes devraient être réalisées dans cet ordre. Cependant, deux étapes consécutives devraient se chevaucher dans le temps de manière substantielle.

Constituer une équipe

Ce projet nécessitera une coopération étroite entre différentes personnes :

- un chef de projet,
- un comité scientifique international,
- des chercheurs (en particulier pour la partie 3),
- un spécialiste en bases de données sur les sciences de la paix,
- des personnes-ressources (en particulier, les administrateurs des diverses séries de données),
- un programmeur,
- un spécialiste des questions de licences logicielles et de copyright,
- des testeurs (pour tester l'ergonomie, le bon fonctionnement pour diverses plateformes, navigateurs et lieux de connexion...),
- la communauté des sciences de la paix.

Certains des membres de l'équipe pourraient avoir plusieurs fonctions, selon les ressources et les compétences disponibles.

Établir des partenariats

En plus des partenaires représentés dans l'équipe du projet, des partenariats spécifiques devraient être établis dans les domaines suivants :

- financement ;
 - fournisseurs de données (universités et centres de recherches, institutions internationales, ONG...).
- La question des données sous copyright est cruciale. En outre, autant que possible, ces partenariats incluront l'aide d'une personne-ressource pour chaque source de données, facilitant la résolution des problèmes liés à la fusion des données (par exemple en cas de documentation ambiguë).
- hébergement internet (par exemple par une université ou un centre de recherche).

Définir précisément les spécifications du projet

Cette étape implique de définir un noyau de fonctionnalités (parties 1 et 2) et le contour précis du programme de recherche (partie 3), de même qu'un objectif en termes d'échéances.

Pour les parties 1 et 2, cette étape nécessitera de définir un ensemble de fonctionnalités souhaitées pour la version 1 du système d'information (dont la liste des bases de données fusionnées). Ces

éléments dépendront des points suivants :

- fonctionnalités et données souhaitées pour la partie 3,
- souhaits et conseils exprimés par les membres de la communauté en sciences de la paix (un questionnaire pourrait être diffusé),
- facilité et faisabilité techniques,
- financement.

Les contours précis du programme de recherche (partie 3) devraient être définis par les chercheurs en accord avec le comité scientifique et le chef de projet. Les données et fonctionnalités requises pour la partie 3 devraient être incluses dans les parties 1 et 2.

Comme proposition provisoire d'échéance, on peut souhaiter disposer de la version 1 du système d'information **6 à 10 mois** après le début du projet (en fonction des spécifications et des ressources disponibles). Le système pourrait être constitué de tout ou partie des bases de données mentionnées ci-dessus, et de certaines des fonctionnalités (niveau « découverte » et l'essentiel du niveau « standard » si possible).

Choisir un cadre de développement

Définir le cadre de développement (c'est-à-dire le(s) langage(s) de programmation, le(s) logiciel(s), l'EDI, les librairies...) du système d'information nécessitera de conduire une analyse comparative approfondie des alternatives techniques disponibles. Une étude préliminaire a déjà commencé ; elle montre qu'il existe plusieurs cadres de développement prometteurs, certains d'entre eux étant entièrement basés sur des logiciels libres (*open source*) et gratuits. Un aperçu de leurs possibilités est montré dans le document joint (Annexe 1). Il est ici suggéré que le système d'information n'utilise que des logiciels libres *et* gratuits, voire qu'il soit publié sous une licence *open source*, par exemple GNU GPL ou CeCILL (cela ne présuppose *pas* de publier les données elles-mêmes sous une licence libre telle que la GFDL).