



Conférence interdisciplinaire
Extraction, traitement et visualisation de données complexes en géographie
(XVIII^e siècle - XIX^e siècle)

Session 3 - Systèmes d'information géographique, Géomatique et visualisation des données.

Séance du 15 mars 2022, 14h – 16h 30

Ecole Normale Supérieure de Lyon, bâtiment Recherche (D4), salle D 024 (rez-de-chaussée) et en ligne
Inscription (gratuite mais obligatoire) : le formulaire est disponible sur le site internet du séminaire
: <https://gitlab.liris.cnrs.fr/geode/seminaires-ixxi>

Programme

1. Présentation de la thématique 3 : Systèmes d'information géographique, Géomatique et visualisation des données.
2. Communications présentées :

T. Joliveau (U. St-Etienne/EVS)

Inventaire, localisation et cartographie des lieux cités dans l'Encyclopédie de Diderot et d'Alembert.

Thierry Joliveau, Université Jean Monnet de Saint-Etienne, UMR CNRS Environnement-Ville-Société

Dans le cadre du projet interdisciplinaire GEODE (<https://geode-project.github.io>) nous nous intéressons à comprendre comment l'Encyclopédie rend compte de la science géographique et des connaissances géographiques de son époque. Un volet de ce travail est de rendre compte de la manière dont les Encyclopédistes nomment, définissent, classent, décrivent et situent les objets géographiques dont ils parlent. La localisation des articles consacrés spécifiquement aux lieux par l'Encyclopédie de Diderot et d'Alembert est évidemment une question intéressante. Quels sont les lieux que cite l'Encyclopédie et ceux qu'elle ignore ? Le mode de la localisation et sa précision varient-ils en fonction de l'espace géographique, le continent par exemple ? Dans l'Encyclopédie, la localisation se fait soit de manière absolue, par des coordonnées en latitude et longitude, soit de manière relative en précisant une distance et une direction ou une relation topologique à un autre lieu.

La communication se focalisera sur un travail en cours consacré à l'extraction automatique et semi-automatique des coordonnées géographiques associées aux lieux dans le texte et aux difficultés qu'elle soulève. Une fois les coordonnées extraites, l'analyse de leur positionnement pose des problèmes spécifiques liés à leur incertitude et leur précision, dépendant des conditions des mesures de l'époque et en particulier au choix du méridien d'origine ainsi qu'aux sources des données. Les coordonnées sont souvent issues de la compilation d'ouvrages de nature différente, plus ou moins précis et plus ou moins récents. La comparaison des coordonnées de l'époque avec les coordonnées actuelles est une piste intéressante mais qui demande de réaliser une correspondance la plus automatique possible entre les toponymes de l'époque et les toponymes actuels. Par ailleurs, une majorité des articles ne comportent pas de coordonnées et ne sont situées que relativement à d'autres lieux, ce qui demande un autre type d'analyse textuelle. Enfin certains lieux sont cités dans articles sans en constituer la vedette et une analyse exhaustive suppose de les extraire sous forme d'entités nommées et de les relier aux articles qui les citent.

Ce premier travail exploratoire ouvre le chantier d'un projet combinant Système d'Information Géographique et Traitement Automatique du Langage Naturel visant à constituer deux objets "manquants" à l'ouvrage : l'index géographique et la carte des lieux de l'Encyclopédie. Il s'agit de mobiliser des méthodes modernes pour en quelque sorte continuer l'Encyclopédie et répondre aux objectifs décrits par Diderot dans son fameux article "Encyclopédie" qui conseillait dans ce domaine de s'en tenir "à la seule connoissance géographique des villes qui fût scientifique, à la seule qui nous suffiroit pour construire de bonnes cartes des tems anciens, si nous l'avions, & qui suffira à la postérité pour construire de bonnes cartes de nos tems, si nous la lui transmettons ;"

Alexis Litvine (U. of Cambridge/CAMPOP)

The COMMUNES project: a geospatial data infrastructure for France 1750 to the present (A. Litvine, T. Thévenin, I. Séguy)

Establishing how many people inhabited a particular area and being able to define precisely its boundaries is not just an academic exercise. It is the cornerstone of historical, economic, and social scientific research. This exercise comes with several challenges, not least the significant landscape and boundary changes over the last two centuries.

This session will cover the background and preliminary results of the *COMMUNES* project, which aims to do this in the long run for France. Combining modern GIS techniques and large historical databases it is now possible to represent and analyse the spatial dimension of demographic and socioeconomic phenomena at an unprecedented level of accuracy. Municipal-level Historical GIS (H-GIS) data now exist for several countries, but France, despite the existence of all the necessary records, is clearly lagging in this respect. The COMMUNES project aims: **i)** to produce the first H-GIS providing annualised boundary data for all French municipalities from the Revolution to the present day; **ii)** to link each spatial unit to historical population data and metrics of access to networks of transport; **iii)** to develop a multimodal model of transport to allow the dynamic analysis of the relationship between economic and demographic changes and accessibility from the end of the 18th century to the present day; **iv)** to represent and analyse on a very fine scale some key demographic and socio-economic phenomena. In this paper we will present our method for reconstructing historic administrative boundaries combining the study of historical and cartographic sources, as well as other yet unexplored documents from local administrative archives and cadastral surveys and link them to existing census data. I will also introduce recent exploratory work aimed at creating population estimates for ancien régime parishes for the early 1780s. In this talk I will **i)** present our method to link ancient parishes to post-revolutionary communes, **ii)** present the historical sources used in this work, and **iii)** discuss the usefulness and potential pitfalls associated to using BMD data to reconstruct population estimates.

I will finally briefly outline the intellectual agenda of the project, including:

Producing high-resolution, spatially disaggregated data for economic history and historical demography is not adding more data to an already crowded field. It opens new ways to answer fundamental questions: it makes possible for the first time real multi-scalar analyses ranging from local to transnational levels and permits the identification of specific characteristics and spatial linkages with adjacent areas, which are central to understanding the diverse demographic and economic trajectories of regions.

Reviving economic geography through spatially linked population, occupational and environmental data. Today, the scale of a region and the spatial relationships between economic activities, infrastructure and institutions are widely understood to be essential to the course of economic development. Yet nearly all existing historical information derives either from extremely localised case studies or has been estimated at the ‘national’ level on the basis of non-standardised proxies. This project brings together demographers, historians, geographers, spatial analysts, and economists to provide a solid empirical bases for historical study.

French economic development through population density: We explore reconstructing and mapping population density at a very fine level as a proxy for early shifts in occupational structure. Spatially disaggregated population data are crucial to the aims of our project: **i)** population data are typically available with more complete spatial coverage than most other types of data, are often available at finer levels of spatial disaggregation, and with higher frequency. **ii)** The distribution of population fundamentally reflects the spatial patterning of economic activities. **iii)** Population data is required as a denominator for the calculation of sectoral shares, participation rates, and as a check on the reliability of occupational datasets. Without precise population data all these other metrics are based on questionable assumptions. **iv)** Population data provide direct evidence of urbanisation, and indirect evidence of migration patterns. **v)** Population counts can be used to inform the collection of other types of data, where budgets are limited. For example, information on population geography can provide a basis for rational sampling for occupational data collection, as it can be digitised for a fraction of the cost of occupational data given the relative simplicity of the information required.

The structuring effect of transportation networks: Our data will offer major new opportunities to analyse the relationship between space, transport network, patterns of population distribution and economic development. The “structuring effect” of transport infrastructures on space is a popular theme in the literature yet it remains mostly an unverified phenomenon. Our geo-historical approach will allow us to both verify the role of population density on the architecture of transport networks and, conversely, the role of transportation infrastructure on population dynamics, and quantify rigorously the substitution effects from road and waterways to the railways.