MUNICIPALITIES, MERGERS, AND THE OUTWARD EXPANSION OF THE CITY

Andrew Sancton
University of Western Ontario

ABSTRACT
An incremental process of annexation often resolves problems arising from the expansion of urban municipalities’ territories when they are allowed to develop on the fringes of other municipalities. Municipal mergers have been advocated in North America ever since the early nineteenth century. Their proponents, dubbed “consolidationists,” have claimed that municipal mergers promote efficiency by harnessing economies of scale in municipal services, encourage rational land use by facilitating regional planning, favour equity by implementing common tax and service levels within the new boundaries, and foster economic competitiveness. The author suggests that the existence of many municipalities within a given area can and does cause serious problems. Municipal mergers, however, are not the only solution.

RÉSUMÉ
Les problèmes créés par l’expansion du territoire d’une municipalité urbaine qui se développe aux limites du territoire des municipalités voisines sont souvent résolus par un processus graduel d’annexion. Depuis le début du 19e siècle, on a généralement été en faveur des fusions municipales en Amérique du Nord. Les partisans des fusions qu’on appelle les « consolidationistes » prétendent qu’elles favorisent 1) l’efficacité, en produisant des économies d’échelle dans les services municipaux; 2) une utilisation intelligente du territoire facilitant la planification régionale; 3) l’équité, en mettant en place un système de taxation et de services communs à l’ensemble des citoyens du territoire; 4) la compétitivité économique. Depuis la fin des années cinquante, ces déclarations ont été fortement critiquées, surtout aux États-Unis. L’auteur donne à penser que l’existence de plusieurs municipalités dans une même région peut causer de sérieux problèmes auxquels les fusions municipales ne sont pas nécessairement l’unique solution.
Cities make more demands on government than the countryside does. Cities require regulations concerning water-supply systems, sewers, and other public utilities; public spaces; dense networks of roads, and some form of public transit; systems for removing solid waste; and various kinds of special services for people in need who are not supported by their families. Rural areas can manage without such government activities. (For some people this is precisely their attraction; rural life usually means self-sufficiency and lower taxes.) Traditionally, the differences between urban and rural have caused each to have their own institutions of local government. The problem, however, has been that cities expand; often well beyond their official municipal boundaries. How to cope with such problems is the main subject of this essay.

Boundaries of countries and of the constituent units of federations are remarkably stable. Even the break-up of the Soviet Union and of Czechoslovakia did not lead to changes in the boundaries of the pre-existing units, in large measure because most people realized that boundary negotiations would be too difficult, if not dangerous. Such is not the case with municipalities. In many countries municipal boundaries change frequently (Anderson, 1996). But even where they do not, there are usually frequent proposals for change, or at least for the creation of new inter-municipal institutions, such as special-purpose bodies for the provision of specified services to a number of neighbouring urban municipalities.

International boundaries are rooted in history and are rightly seen as having great political, economic, and social significance. Municipal boundaries in most places were designed for functional purposes. Some people at some time thought they appropriately designated an area that was suited for its own local council and, consequently, a common package of local public services. Such functional considerations are most obvious when we examine the original boundaries of urban municipalities: cities, towns, and villages. It was usually precisely because these areas became urban centers requiring different public services from the surrounding countryside that they became established as municipalities in the first place. The original boundaries of rural municipalities are less obviously functional, although they often relate to what were considered reasonable distances for travel at the time they were first drawn. In some European countries they survive from feudal times.

Because municipal boundaries, especially urban ones, were originally drawn for functional reasons, they are invariably the subject of continuing debate (Sancton, 2000). The claim that certain munic-
Principal boundaries are outmoded or dysfunctional makes much more political sense than a similar claim about a country’s boundaries. Calls for change in municipal boundaries are usually caused by the outward expansion of urbanization beyond an area officially designated as a city, town, or village. Sometimes, proposals for change are genuinely motivated by a desire for a more effective and fair way of making local decisions; on other occasions some participants in a conflict about municipal boundaries are maneuvering for political or economic advantage.

Annexation and its problems

Let us assume that sometime in the past an area of urban settlement is officially designated by the central government as a town. A town council is elected to oversee the provision of local public services. The boundaries include all the territory containing relatively dense (non-rural) residential settlement and all the factories, shops, and offices in which the residents of the town are employed. Because the authors of the boundaries are far-sighted, they even include within the town some land to accommodate future urban growth. Let us also assume that town residents are generally successful in their economic endeavours: businesses grow and more people move to the town because there are good employment opportunities, and because the council has helped insure that it is a pleasant place to live.

Sooner or later, there is going to be a problem: the town will have no more land for new development. There are three possibilities. The first possibility—the most unlikely—might be that new development simply cannot happen: urban development in rural areas might not be allowed and town residents might accept a halt to growth (for the short-term at least, the values of their own properties will rise because they possess a scarce commodity, urban land).

The second possibility—much more likely—is that the council of the surrounding rural municipality will be willing and able to approve urban growth, though without a change in boundaries. Depending on arrangements for the financing of municipal services, such a policy might have the effect of lowering taxation levels for current rural residents, because the resulting population growth in the rural municipality means that they would now have more people with whom to share costs. Rural residents owning property near the town would be especially fortunate because their property values would inevitably increase as the official designation of their land changes from rural to urban/industrial use. Even new urban-minded-
ed residents in the rural municipality might consider themselves better off than if they had located in the town. They might be able to escape paying for every facilities and services (that is, the ones located in the adjacent urban centre) that were a factor in causing them to want to move to the area. This second possibility contains the seeds of many of the issues and problems to be discussed in the rest of this essay.

The third possibility is that the boundaries of the town be extended so as to accommodate the new growth. This process of boundary extension is known as “annexation.” Municipal annexation is rarely easy, but it is less difficult when the land in question is rural rather than already being fully developed. Since the most obvious purpose of annexation is to allow for a single urban municipality to provide for orderly outward urban expansion, it makes sense that, ideally at least, annexation should occur prior to development rather than after.

If annexation is to be allowed at all, central governments must provide a procedure for it. Such a procedure might involve the necessity of some form of local negotiation and agreement. The agreement could involve only the two councils or there might be a requirement for approval by various local referenda, especially involving those residents and property-owners most directly affected by the proposed change. The problem with requiring local agreement is that there is rarely any incentive for the municipality that is losing land to agree. Compensation payments, however, can sometimes be used to purchase agreement. The position adopted by landowners in the affected area usually depends on whether or not annexation will increase the value of their land. If development is only possible if the land is within the boundaries of the central urban municipality, then they will likely be in favour. If development is possible whatever the outcome of the annexation dispute, then their position will depend on taxation levels, and particular local political circumstances.

An alternative to local agreement is for the central government to allow annexation issues to be settled by the courts, that is, by some form of administrative or quasi-judicial tribunal, or by some special authority created only to deal with municipal boundary disputes. The problems here are predictable. Such a process can be very expensive, especially when highly paid lawyers and experts become involved in public hearings. Sometimes, however, the issues at stake are fundamentally political rather than technical, and it is simply inappropriate in a democracy for unelected people (such as those
appointed to sit on tribunals) to make such decisions about the political futures of significant urban areas. This last problem can always be overcome by insuring that the process allows a relevant minister in the central government to overrule the tribunal’s decisions. Such a provision can make life politically difficult for a minister as he or she is pressured to overrule various controversial decisions, while coping with such difficulties, and being accountable for their resolution, is exactly what central government ministers are paid to do.

As long as municipalities are not constitutionally protected against changes in their boundaries without their consent, central governments can always use their legislative authority to sort out boundary disputes. The main difficulty is that most of the members of any central legislature will, almost by definition, have no interest in a local issue that does not affect their own constituencies. At best, such legislation is a waste of their time; at worst it is an opportunity for them to extract favours from one side or another in return for their vote. Only when the resolution of a boundary dispute creates significant precedents, or when it affects the future of the very largest of cities, would such an issue seem appropriate for legislative settlement.

Regardless of how they are brought about, annexations, while solving some problems, can create new ones of their own. For example, even if everyone accepts in principle the need for an annexation, there can be serious disagreements about its size. Should there periodically be huge annexations or is it better to take in a few hectares on a regular basis? What happens when so much of a rural municipality has been annexed that the remainder is simply not viable? Should the urban municipality take everything or should the remaining rural area be attached to other rural municipalities? What if no municipality wants to absorb the rural remnant of an annexation?

Sometimes the threat of an unwanted annexation provokes residents or landowners in a rural area to try to establish a new urban municipality within the disputed territory, the belief being that it will be more difficult (legally or politically) for an entire urban municipality to be annexed than part of a rural one. Whether or not this calculation is correct, and whether or not such new incorporations are possible, depends greatly on the particular legal regime established by the central government. In the United States the calculation has generally been correct and the incorporation of new municipalities is relatively easy. This helps explain why many large American metropolitan areas contain dozens, if not hundreds, of dis-
tinct urban municipalities (Burns, 1994).

Another problem with annexation results when two nearby urban municipalities each absorb so much previously rural land around their original boundaries that they eventually become contiguous jurisdictions, each responsible for part of what has become a single urban area. In such circumstances one fact is clear: further annexation along the two municipalities’ common boundary is difficult, if not pointless. Indeed, for many people, whom we can label as “consolidationists”, it is the boundary itself (which now separates two adjacent municipalities within a single built-up area) that has become the problem.

The case for consolidation
As the previous discussion has shown, it is quite possible for a single urban area to evolve in such a way that it contains many separate municipalities. Adjoining rural municipalities might have become urban over time; new urban municipalities might have been established within a former part of a rural municipality; two previously distinct urban areas might have grown into each other; or all or some of these processes might have occurred simultaneously. Ever since the mid-nineteenth century there have been politicians, civil servants, academics, and sometimes even real-estate developers who have observed such phenomena with growing concern. They began to refer to the increasing number of municipalities within a single urban area as “fragmentation.”

Consolidationists are opponents of fragmentation (Rusk, 1995). Over the past one hundred years at least they have powerfully stated their case. Although it can be summarized in different ways, the consolidationists’ argument effectively comprises four main points. Each will be described in turn.

The first is that fragmentation is inefficient. Consolidationists argue that, for some municipal services at least, it will inevitably be more efficient to have a single integrated municipal department covering the entire territory. Such an argument rests on the belief that are economies of scale in municipal services and that they need to be captured.

The second argument relates to urban planning. Here the consolidationists’ belief is that every urban area requires a plan and that a locally-elected council should develop it. The whole purpose of the plan is to lay out in an integrated way how and when present and future systems of roads, public-transit facilities, sewers, utilities, and major public and private buildings will be developed and how they
will relate to each other. Although each separate municipality might have its own limited plan, it is obvious that the sum of such plans does not necessarily mean an integrated, coherent plan for the entire area. Indeed, it is quite possible that the plans will be contradictory, raising the spectre of major arterial roads that end abruptly at municipal boundaries, and of nearby sewage-treatment plants in two separate municipalities, each with surplus capacity.

Consolidationists are also concerned about equity. They believe that citizens of the same central jurisdiction who live in the same urban area should, to the greatest degree possible, benefit from the same level of municipal services and taxes. They are outraged that in some countries, especially the United States, rich people can isolate themselves in their own municipalities, provide themselves with high-levels of service while paying relatively low taxes, all the while leaving the financial burdens of the central city to the poorer taxpayers who are unable to afford expensive houses in suburban enclaves. In these cases, “municipal autonomy” is cited as the reason why it is inappropriate for outsiders to impose change. Consolidationists therefore seek to overcome the barrier of municipal autonomy by merging the central and suburban municipalities into a single entity.

More recently, consolidationists have argued that municipal fragmentation makes it more difficult for an urban area to compete effectively in a globalized economy. There is an assumption in such reasoning that it is the system of municipal government, rather than businesspeople within the city, who are engaged in the competition. To the extent that municipalities engage in advertising to attract investors, it might well be that the absence of a common message on behalf of a single urban area is damaging.

The consolidationists’ greatest victory was more than a hundred years ago, in 1898, when the legislature of New York state merged New York (Manhattan), Brooklyn, Richmond, the Bronx, and Staten Island to form a new City of New York. Since then, the city has indeed been a remarkable economic success. None of its historians, however, have ever attributed such success to consolidation. A more common claim is that the parts of the city outside Manhattan have been worse off as a result of the consolidation than they would have been had they retained their municipal independence. In any event, if the New York area could be considered consolidated a hundred years ago, it scarcely can be now. The City of New York now includes only 7.3 million of the 18.1 million people that the American federal Bureau of Statistics considers to be part of what it calls the Con-
solidated Metropolitan Statistical Area of New York (Sancton, 1998b, 2–3). The city’s municipal boundaries have not changed since 1898 but growth continues all around it (Danielson and Doig, 1982).

One of the consolidationists’ main difficulties has been agreeing on appropriate boundaries. For the purposes of regional planning, boundaries should probably extend quite far out into the countryside, especially if there is continuing pressure to allow people who work in the city to live and build wherever they want in the countryside. But, for other consolidationist objectives, such boundaries might take in too much. For example, the objective of efficiency is compromised if such services as garbage collection and public transit are extended into sparsely-populated areas where unit costs are high. Similarly, the equity objective is weakened if farmers are forced to pay for city services they rarely use.

A longstanding consolidationist answer to such concerns is to propose a two-tier system of urban (or metropolitan) government. For many consolidationists such a system is superior to complete consolidation, especially for very large urban areas. A two-tier system of municipal government is one in which one municipal council is established to cover the entire urban area, but local councils are retained in order to manage those municipal functions that seem more local in scope. In short, a two-tier system of municipal government is like a municipal federation, with the central government determining the nature of the federal constitution, or at least acting as a referee between the two levels.

The best known urban two-tier local-government systems have been in London and Toronto. Although the upper-tier London County Council (LCC) was established in 1888, the 28 relatively weak lower-tier “metropolitan boroughs” were not organized until ten years later. In 1965 the British parliament abolished the LCC and replaced it with the Greater London Council (GLC). The GLC, comprising 32 “Greater London boroughs” that were functionally stronger than their lower-tier predecessors in the LCC, covered a much wider territory. The Labour-dominated GLC was itself abolished in 1986 when Margaret Thatcher, a conservative, was prime minister. In 2000, the Labour government, led by Prime Minister Tony Blair established a new upper-tier council for metropolitan London, complete with Britain’s first directly-elected mayor. Now that the new council is in place, there is once again a local authority capable of making strategic infrastructure decisions for the London area. More importantly, for the first time ever, London will have its own powerful political spokesperson: a directly elected mayor.
The implications for Britain’s domestic political process are difficult to predict, but it would be surprising indeed if they were not significant.

Toronto’s upper-tier local government, the Municipality of Metropolitan Toronto (Metro) came into being in 1954, bringing together thirteen different municipalities. These were consolidated into six in 1966. However, in 1998, Metro and its six constituent units were all merged into the new city of Toronto (Frisken 1998). A much weaker upper-tier authority, the Greater Toronto Services Board, covering a greatly-enlarged territory, began operations in 1999. Experience in London and Toronto indicates that two-tier metropolitan government is at best controversial, at worst, unstable (as will be discussed below). Similarly checkered histories are available for Rotterdam, Copenhagen, Barcelona (Sharpe 1995b), and Winnipeg (Brownstone and Plunkett, 1983).

Two-tier systems only partially solve boundary problems. Few people are ever satisfied with the upper-tier boundary: for land-use planners the territory is usually too small; rural or outer-suburban residents have difficulty understanding why they must be included if they rarely go near the city centre. The lower-tier boundaries can also cause difficulties, especially if the representational system at the upper tier treats all lower-tier units with a degree of equality, by entitling each to the same number of representatives on the council regardless of their population. But the greatest problem is that, while the upper tier provides at least a partial antidote to geographical fragmentation, the very creation of the upper tier bifurcates the activities of existing municipal governments. Areas previously served by one municipal government are now served by two. Coordinating their activities within dense urban settlements is possible, but not easy (Sharpe 1995a).

What seems to be almost impossible is working out acceptable political arrangements. If there is direct election to the councils of both levels and political parties are involved (as in Europe), then there is always the danger of inter-party conflicts complicating the relations between the upper tier and at least some of the lower-tier municipalities. If the elections are non-partisan in that no parties are involved (as in Canada), then voters have great difficulty keeping track of individual candidates at both levels; furthermore, the provocation of conflict between the two levels is often an effective mechanism for particular councillors at either level to attract much-needed attention to themselves. If lower-tier councillors choose representatives to the upper tier (as opposed to representa-
tives being directly elected by citizens), then many will argue that the upper-tier council is not properly accountable to the people. The more functions exercised by the upper tier, the greater such an “accountability deficit” is likely to be.

To the extent that consolidationists accept such criticisms of two-tier systems, they tend to advocate complete single-tier consolidation as the alternative. But in the case of large urban areas, they are therefore once again faced with their original difficulty: different local-government functions can most effectively be provided at different population levels. Outright consolidation creates as many problems as it solves. As the second half of the twentieth century progressed, it seemed to many that the case for consolidation as a solution to problems of urban governance was intellectually bankrupt.

The public-choice alternative

Consolidationists tended to look at the problems of urban governance from the perspective of political science and public administration. For them, problems of inefficient or ineffective urban governance were usually caused by municipal governments not being big enough. When economists looked at the same problems, they often saw big municipal governments as the problem, not the solution. The “public-choice” approach to the organization of local government in urban areas stems from economics and is aimed at harnessing the discipline of the marketplace to the provision of local public services. Public-choice analysis emerged in the USA in the late 1950s and 1960s. By 1987 the American federal government’s Advisory Council on Intergovernmental Relations had endorsed the public-choice approach and rescinded its earlier support for consolidationist policies (Keating, 1991, 111). Although not as influential in other countries, public-choice analysis must be taken seriously in any attempt to prescribe ideal arrangements for the organization of local government in large urban areas.

Advocates of public choice refute the main claims of the consolidationists (Oakerson, 1999). They acknowledge that, for some local-government services, large territories are required to capture economies of scale. But they provide ample evidence to demonstrate that other services are best provided over smaller areas. In their analysis, they break down most services into their various parts. Policing, for example, consists of foot patrol, crime detection, dispatch, training, and a host of other activities. Each such activity likely has a different optimum scale of population. Better to keep municipal units relatively small, give them initial responsibility for
all local-government functions, and anticipate that they will cooperate with their neighbours for matters that require larger territories. In some cases, cooperation might involve the establishment of special purpose bodies to provide a particular service over a long period of time for a specified multi-municipal territory.

The main argument of the public-choice approach is that many municipalities within the same urban area should be maintained so that they will be forced to compete with each other with respect to taxation levels and quality of services. Some might offer specialized bundles of taxes and services that are attractive to particular residents or business people. Just as competition encourages efficiency in the supply of private goods, advocates of public choice suggest that competition among nearby municipalities produces similar benefits. This is how supporters of the public-choice approach counter the argument that the consolidation of municipalities increases efficiency.

Public-choice analysis begins from the assumption that governments cannot predict the future. Consequently, it places little value on the capacity of any local government to produce a useful regional plan. Drawing from the conceptual toolbox of economics, public choice holds that most participants involved in the process will be “rent-seeking,” i.e. they will be concerned with protecting their own narrow interests rather than seeking a mythical “public interest.” Politicians will be trying to find ways to reward their friends and supporters; staff planners will be looking out for their own jobs, and proposing new projects to advance their bureaucratic empires; consultants will be generating potential new contracts for themselves; owners of existing buildings will want the plan to protect the value of their properties; owners of undeveloped land will want freedom to develop. According to public choice, creating the potential for an enforceable regional plan simply creates more opportunities for the rent-seekers without necessarily improving the quality of the built environment. Pointing to infrastructure development in American cities that have never had formal regional plans, public-choice advocates claim that various and complex networks of municipalities, special-purpose bodies, central-government agencies, and private contractors are able to plan, finance, and build large projects in an orderly, coordinated fashion. Formal plans, they claim, are not necessary. Critics of this argument suggest that the automobile-centred characteristics of American cities are evidence of poor or non-existent planning. They are in turn countered by the claim that this is exactly what most Americans want—for most—the system works.

What of the citizens who do not get what they want, who cannot
afford to live in the suburbs or own a car? What should the local government system be expected to do for them? For these difficult questions, public-choice has a quite simple answer: nothing. The public-choice position is that any decisions about redistributing income from rich to poor should be made at the central-government level because local governments cannot at the same time compete effectively with each other with respect to taxes and services and be expected to help poor people. If central governments redistribute income, then poor people have more freedom in deciding where to live and how to use their own resources. If they (and others) want good public transit, then it will be economically viable. If, however, poor people would prefer to travel by personal automobile, then they should be able to decide how much of their own scarce resources they want to use to attain this goal. Critics of the public-choice approach to local government point out that, with increasing globalization, even central governments must compete with each other to attract investment by offering high-quality services and low taxes. Under these conditions, central governments face the same economic constraints relating to income redistribution as local governments do. Assuming income redistribution is desirable, the critics of public-choice simply claim that large local governments are more likely to do it than small ones.

Advocates of public choice believe that municipalities compete with each other with tax and service levels, not with promotional brochures, media advertising, and trade missions. If the public-choice approach is correct, municipalities need not be consolidated in order to improve their opportunities for economic development. Intermunicipal public-private cooperation is desirable with respect to whatever marketing directed at outside investors is necessary, but consolidation and economic development, from a public-choice perspective, are quite unrelated, except for the possibility that large, consolidated municipalities might be so inefficient that they harm the competitive position of their respective urban areas.

Conclusion
It is difficult to find much common ground in the consolidationist and public-choice positions. There are, however, two points of agreement: 1) for some local government functions at least, there are needs to some form of local institution capable of operating over an entire continuously built-up urban area; and 2) small municipalities within large urban areas should not be responsible for deciding on policies involving the redistribution of income from the rich to the
poor. Almost everything else is in dispute. This means at a minimum that there is no reliable “expert view” about how municipal government should be organized within urban areas. The great advantage of such uncertainty is that there is no compelling reason to rush any kind of structural reorganization. Instead, people in local government can concentrate on improving the operation of existing institutions rather than fantasizing about the great improvements that would hypothetically result if only boundaries and functions could be changed. The other advantage is that officials can listen to the demands of ordinary people. Whether they favour many small municipal governments and a network of special-purpose bodies, or one large municipal government, or a two-tier system of some kind, there is a strong expert case to be made for whatever arrangements local citizens prefer.

NOTES
1 For a case study treating most of the issues examined here, see Sancton, 1998a.
2 For a review of the literature concerning the relationship between costs and the size of municipalities, see Boyne, 1995.
3 For critiques of this approach, see Lyons, Lowery, and deHoog, 1992 and Keating, 1995.

REFERENCES
Berkeley: Institute of Governmental Studies Press at the University of California.


*Editor’s Note: This paper reflects some of Professor Sancton’s comments in his keynote address at the ETRC colloquium. It is also based on a similar essay he published, titled “Metropolitan and Regional Governance” in Fowler, E.P and Siegel, D. (eds) Urban Policy Issues: Canadian Perspectives, 2nd edition 2001. Toronto: Oxford University Press.*