

# Intelligent Middleware for Understanding Neighborhood Markets

## Proposal

submitted to

### The Urban Markets Initiative of the Brookings Institution

by

Prof. Joseph Ferreira, Jr.  
**Massachusetts Institute of  
Technology**

and

Marc Draisen, Executive Director  
**Metropolitan Area Planning  
Council**

April 22, 2004

#### Contact Information:

MIT, Dept. of Urban Studies and Planning	Metropolitan Area Planning Council
<b>Joseph Ferreira, Jr.</b> Professor of Urban Planning and Operations Research MIT Room 9-532 77 Mass Ave, Cambridge, MA 02139 Tel: 617-253-7410 Fax: 617-253-3625 Email: <a href="mailto:jf@mit.edu">jf@mit.edu</a>	<b>Holly St. Clair</b> Director, MAPC Data Center Metropolitan Area Planning Council 60 Temple Place Boston, Massachusetts 02111 Tel: 617-451-2770, x2014 Email: <a href="mailto:hstclair@mapc.org">hstclair@mapc.org</a>

# Intelligent Middleware for Understanding Neighborhood Markets

Submitted to: **Brookings Institution, Urban Markets Initiative**

Date: April 22, 2004

From: **Prof. Joseph Ferreira, Jr.,** & **Marc D. Draisén, Executive Director,**  
Massachusetts Institute of Technology Metropolitan Area Planning Council

## Executive Summary

**Project Lead:** Joseph Ferreira, MIT Room 9-532, 77 Mass Ave, Cambridge, MA 02139 Tel: 617-253-7410, Fax: 617-253-3625, email: [jf@mit.edu](mailto:jf@mit.edu)

### **Collaborators:**

Holly St. Clair, Director, MAPC Data Center, Metropolitan Area Planning Council

Larry Braman, Manager, Mapping and Data Services, Boston Dept. of Neighborhood Development

Charlotte Kahn, Director, The Boston Community Building Network, The Boston Foundation.

### **Goals, Objectives:**

The proposed collaboration will prototype and test an intelligent middleware approach for sharing data within a metropolitan area in a manner that is likely to be more effective, scalable, and sustainable than the traditional 'data center' approach. The proposed tools and methods also provide a mechanism for accumulating local knowledge about neighborhood-scale land use, ownership, and market potential and for using that knowledge to re-interpret administrative datasets and develop customized analyses of neighborhood conditions and market potential. The collaborators include local and regional agencies with in-house GIS expertise and significant experience acting as 'data centers'. These agencies are ready and willing to collaborate with MIT in the development and testing of data-intermediary approaches that can facilitate both horizontal data sharing across agencies and towns and vertical data sharing among regional, local, and neighborhood organizations. They have been working for several years with the Boston Foundation as local partners in the National Neighborhood Indicator Partnership and they are interested in methods that can enhance the reusability of administrative data and streamline its integration with locally generated neighborhood data, to permit customization of neighborhood indicators and analyses.

### **Methodology:**

The project develops and tests prototype tools and methods for sharing and augmenting the data needed to understand neighborhood markets and community development options. University expertise is tapped to design next-generation tools for data sharing. University researchers, local and regional agencies, and community organizations collaborate closely in prototype development and in several workshop and class projects that test the practicality of the new methods. Advisory Board input and continued involvement in related national efforts provide useful input and several outreach and communication channels.

### **Impact**

The project team has been extensively involved in efforts to improve the quality and utility of data sharing efforts that can inform community development and metropolitan planning. The team includes Boston's NNIP partners and has been active in recent efforts to establish a nationwide Community Statistical System Network. A number of planners and community organizations across the country have expressed interest in finding alternatives to the traditional 'data center' model and are interested in implementing the middleware tools and methods developed by the project if the prototype proves effective.

# Intelligent Middleware for Understanding Neighborhood Markets

## I. Introduction and Needs Assessment

As evidenced at the recent Community Indicators Conference<sup>1</sup>, a large and growing number of local agencies and community organizations are actively engaged in efforts to quantify and evaluate the social, economic, and cultural health of their neighborhoods. Increasingly, these efforts involve the use of city, regional, and federal data, together with various GIS and statistical tools. They have been fueled by the rapid growth and computerization of detailed, spatially disaggregated databases, together with improvements in software, distributed computing, and related information and communication technologies. These efforts make it possible to quantify, chart, and map various community indicators in order to assist in understanding neighborhood-scale market conditions, forecasting programmatic needs, and providing decision support.

While these efforts can be quite helpful, the more useful ones tend to be very labor-intensive and not easily replicated or sustained. Even if the indicators are developed primarily from existing administrative data, rather than from customized, special-purpose surveys, the development of meaningful small-area indicators generally requires complex data integration and statistical analysis. Often, administrative data needs to be reinterpreted and cross-referenced with local data before its meaning can be ascertained.

The prevailing model for handling these data integration and analysis steps is to rely on ‘data centers.’ Many of the more advanced local efforts are in metropolitan areas that are part of the National Neighborhood Indicators Partnership<sup>2</sup>. Typically, one member of the local partnership will take the lead in establishing a ‘data center’ that acquires copies of relevant datasets from many other local agencies and organizations, integrates them into one system or ‘data warehouse,’ constructs a set of indicators and statistical reports, and disseminates the results through various online tools and printed reports. This ‘data center’ model is depicted in Figure 1-A. Each data center must undertake all three steps: acquiring the datasets, doing the data processing needed to integrate and interpret the datasets, and constructing the data services that deliver the targeted analyses and reports to their constituents. As the number of relevant data sources increases and the desired indicators and analyses become more specialized and complex, the data centers are getting bogged down and/or splinter into isolated data centers with different specialties.

Some of the more sophisticated data centers are recognizing these limitations and trying to become more ‘virtual’ – see Figure 1-B. In this case, they try to become specialized ‘portals’ that provide a window into an assortment of ‘backend’ data repositories without having to duplicate and fully integrate all the individual datasets. Generally, they begin by trying to work with each data repository agency to standardize the flow of information to the data center and make the data sharing more routine. Typically, this involves agreeing to a standard format for a ‘snapshot’ of repository data that summarizes the monthly or annual status of sharable data. Such a dataset, for example, might be summary records for each parcel of land that indicate the current characteristics, land use, ownership, and assessed value of each parcel. While the ‘virtual’ data center approach is an improvement, the data integration and interpretation tasks at the ‘portal’ end are still labor-intensive and hard to maintain. The ‘portal’ terminology suggests a straightforward ‘window’ into the data repositories that helps the user pick the relevant datasets and packages the results in a familiar format. However, actionable knowledge that can contribute to effective market analysis and planning at a neighborhood level is not readily automated as a linear sequence of simple data-processing steps. Much more work needs to be done to develop intelligent data intermediaries that can accumulate and reuse the

---

<sup>1</sup> “The 2004 Community Indicators Conference: Advances in the Science and Practice of Community Indicators,” Reno, Nevada, March 10-13, 2004; <http://business.wm.edu/isqols/community>.

<sup>2</sup> The Urban Institute website for the National Neighborhood Indicators Partnership is available at: <http://www.urban.org/nnip>

local knowledge needed to cross-reference, augment, and re-interpret the ‘official’ datasets. Figure 2 illustrates the role of such ‘smart’ *middleware* in developing the analyses and indicators that are needed. Even if we standardize the format and meaning of the dataset ‘snapshots’ and the end-user data services, translating between the two involves complex data processing. This is especially true if we want the interpretation of the repository data to depend, in sophisticated ways, on local knowledge supplied by the end users. Yet, many of the data processing, analysis, and interpretation steps can be anticipated and pre-packaged in ways that can be standardized and tuned to the needs of community groups and local agencies. Focusing on such ‘intelligent data intermediaries’ can help streamline data-sharing efforts while greatly enhancing their likelihood of empowering grassroots planning and of nurturing urban markets. Developing such tools requires a significant paradigm shift away from data centers and data sets and towards distributed data services with interoperable, Lego-block parts and configurable data intermediaries. Before describing our specific project scope and focus, we will highlight a simple, illustrative example of the need for such intelligent middleware.

## II. Illustrative Example – Overcoming ‘Spelling Errors’

To illustrate the types of problems that arise, we use a simple example of administrative data-sharing complexity from a paper, “Information Technologies that Change Relationships between Low-Income Communities and the Public and Non-Profit Agencies that Serve Them,” by one of the researchers<sup>3</sup>. The example involves the use of Boston's parcel database to identify and categorize land-use and ownership patterns. Understanding local land-use and ownership is only one part of an effort to understand a neighborhood’s potential for economic activity. Nevertheless, the example will illustrate the frequent need for, and common frustrations with, integrating and reinterpreting key data sources that are dispersed among various agencies and groups.

Many cities and metropolitan areas have begun to provide annual snapshots of their parcel records – either online or on CDs – for general use by other agencies and, often, the public at large. Such datasets record parcel location, size, assessed value, ownership, and the like and can be quite valuable to many local agencies, organizations, and researchers. For example, we might be interested in identifying the largest landowners in a neighborhood or, perhaps, all city-owned properties that might be targeted for reuse or other disposition.

Public agencies own thousands of parcels out of the 130,000+ parcels in Boston’s official 1996 database. The Boston Redevelopment Authority (BRA) is one of many public agencies that is a big property owner. Within the official city records, there are at least 17 different spellings of ‘Boston Redevelopment Authority’. Among these variations are:

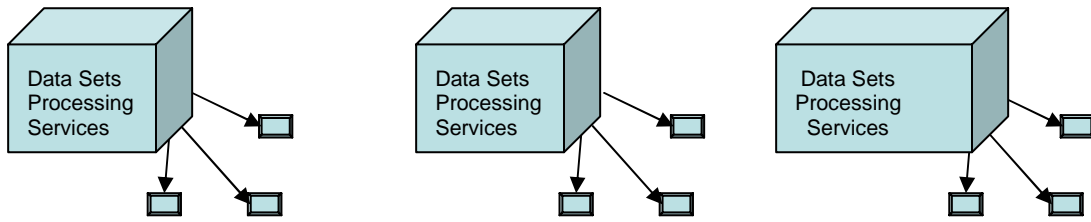
BOSTON REDEVELOPMENT AUTH	231 times
BOSTON REDEVELOPMENT AUTH	1 time
BOSTON REDEVELOPMENTAUTH	83 times
BOSTON REDEVELOPMNT AUTH	82 times
BOSTON REDEVELPMNT AUTH	41 times
BOSTON REDEVELPMNT AUTH	22 times
BOSTON REDEVLPMNT AUTH	28 times
BOSTON REDEVLPMNT AUTHOR	29 times
BOSTON REDVLPMNT AUTH	60 times

---

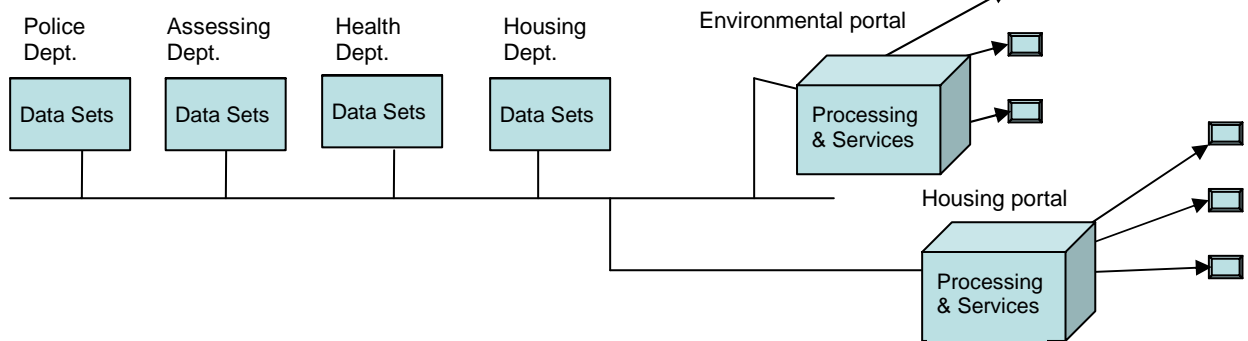
<sup>3</sup> Ferreira, Joseph Jr., “Information Technologies that Change Relationships between Low-Income Communities and the Public and Non-profit Agencies that Serve Them,” Chapter 7 in **High Technology and Low-Income Communities**, D. Schon, et al. (editors), MIT Press, 1998. See: <http://web.mit.edu/sap/www/colloquium96/papers/7ferreira.html>

# Figure 1: Data Center Models

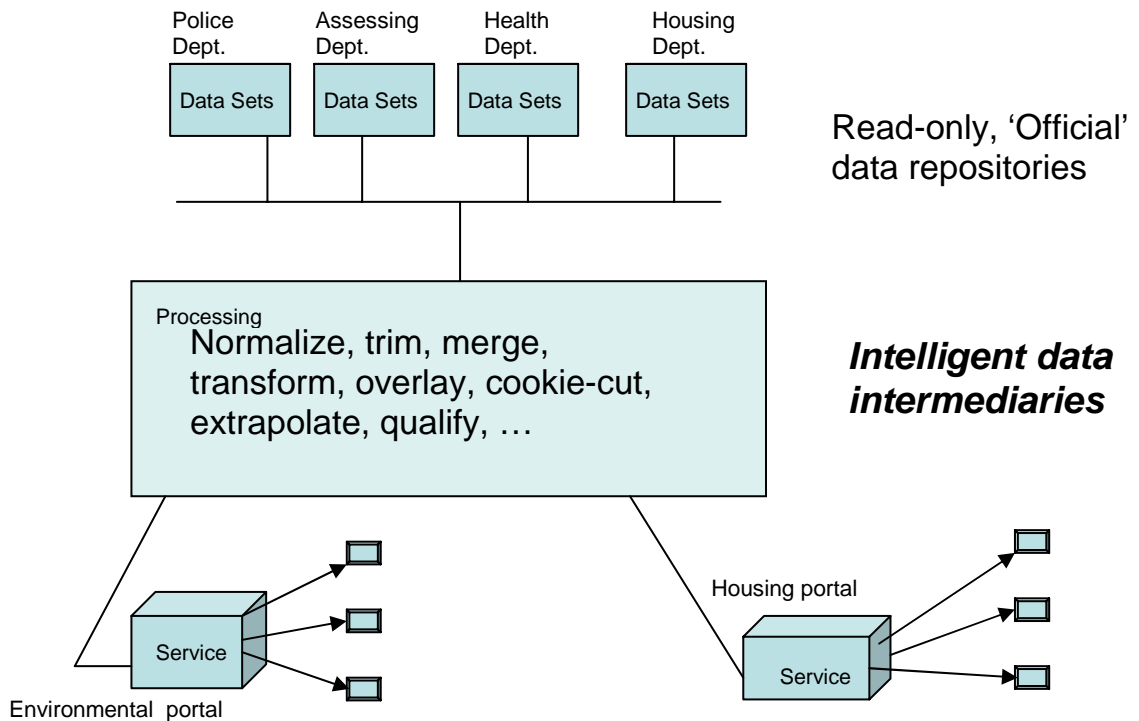
1-A: Isolated data centers (with combined datasets, data processing, and data services)



1-B: Coordinated virtual data centers (with distributed datasets)



# Figure 2: Smart 'Middleware' Alternative



Many other City and State agencies own property in Boston - and many of them also appear as the owner of record with various spellings. Understanding private ownership patterns faces the same problem. For example, there are at least 24 spellings of Boston University (BU) in the parcel database. To facilitate credible analyses of public ownership patterns in a neighborhood, a community data center would have to find and correct such spelling variations. This ‘bottom-up’ approach to fixing data errors is a time consuming but feasible task for one neighborhood, but not easily replicable, and quite inefficient since a new release of the parcel database comes out each year. Alternatively, one might try to catch spelling variations with edit checks at the data entry phase – a tried and true ‘top-down’ information systems approach that tries to trap and correct errors at as early a stage as possible. But the owner of record is determined by the Registry of Deeds, not the Assessing office that maintains and distributes the official parcel records – and the Registry still uses a paper-based system. Moreover, the real issue is not finding and correcting spelling variations but identifying – and recording and saving – the names and relationships among public and private owners whose pattern of ownership is being studied. These relationships are *not necessarily known* at the time of property exchange when official owner names are recorded.

Neither the bottom-up nor the top-down strategy for fixing errors is efficient and practical. Better technologies do exist for handling the problem without any need to change the ‘official’ records. In particular, we can maintain a local lookup table that corrects any misspellings that we find in the official record (and can also categorize each official owner name into our own ownership categories). Whenever we wish to identify ownership patterns in the neighborhood, we can then ‘join’ our local lookup table to the ‘read-only’ official data. We join the tables using the official owner names but we group the owners (to count total acreage or to thematically shade our map) using the *corrected* spellings in our lookup table (or any other locally-defined category of ownership that we store in additional columns of our lookup table). The spelling corrections that we make can be saved in the form of structured query language (SQL) statements such as the two ‘update’ statements in Figure 3. These particular ‘update’ queries search for key syllables in the official owner names in order to identify those parcels that are likely to be owned by the BRA or BU. The ‘update’ statements standardize the spelling of BRA and BU (*in the lookup table*). The original parcel records are not changed – but only reinterpreted by being associated with standardized spellings in the lookup table. Figures 4a and 4b show the breakout of Boston’s largest landowners before and after running the two ‘update’ statements from Figure 3. In each case the same query is run against the parcel table joined to the lookup table. All that has changed between the two queries are the spelling ‘updates’ made to the lookup table. Note that, before running the updates, Boston University does not even make the ‘top-9’ list of Boston owners and BRA’s parcel ownership is significantly undercounted (two-thirds of BRA’s parcels were not counted in the ‘before’ case). After running the updates, BRA and BU become the 3<sup>rd</sup> and 5<sup>th</sup> most frequent owners of Boston parcels.

These technologies (relational database tools and distributed database technology) are increasingly used in business and in advanced data warehouses. The ‘update’ statements can be treated as ‘business rules’ that can be stored and managed within the database and ‘triggered’ to run under specified conditions. Use of such technology provides a means of storing local knowledge about standardized spellings (and more subtle ownership control) so that it can be more easily accumulated and reused for other projects and maintained independently from the official datasets. Lookup tables can be local and private while the parcel data are treated as ‘official’ read-only records maintained by the relevant registry/assessing/planning agencies.

With this ‘middleware’ approach, official databases can be expensive, robust, online-transaction processing systems (Oracle, SQL-Server, etc.) while the lookup tables use inexpensive desktop tools such as MS-Access or can be stored on a local server that performs a data intermediary function. If planners, local agencies, & community groups can count on such a structure, they can use it to develop efficient and effective strategies for local ‘knowledge accumulation’ (e.g., to understand the extent of public influence in a community or to research agglomeration and tax-arrears conditions that might constitute an economic incentive for arson). Modern web services technology and interoperability standards, combined with mature relational database technology allow us to use loosely coupled approaches that permit decentralization while

retaining sufficient structure and standardization to be sustainable. However, access to technology is not sufficient to evolve patterns of use that can take advantage of these. The ‘middle-out’ strategy that we have suggested involves less ‘automation’ and more ‘enabling’ – but it also requires a more sophisticated labor force and attention to learning paths in order to be used effectively. Balancing technical complexity and this need for sophistication is one of the issues to be studied.

**Figure 3: Running UPDATES on the local Lookup Table (called ‘owngroup’)**

```

UPDATE owngroup
SET fixed_owner = 'BOSTON UNIVERSITY'
WHERE oldowner LIKE '%BOSTON%'
    AND oldowner LIKE '%UNIV%';

UPDATE owngroup
SET fixed_owner = 'BOSTON REDEVELOPMENT'
WHERE oldowner LIKE '%BOSTON%'
    AND (oldowner LIKE '%REDEV%' OR oldowner
        LIKE '%REDV%');

```

**Figure 4: Revealed Ownership Patterns Before & After Updates**

Largest landowners - Before 2 UPDATES of the lookup table

OWNER	PARCELS	VALUE (K)
CITY OF BOSTON	2876	3883239
CITY OF BOSTON BY FCL	1258	19117
SANTORO VINCENT J	309	23087
COMMONWEALTH OF MASS	235	528891
BOSTON REDEVELOPMENT AUTH	231	106863
BOSTON HOUSING AUTHORITY	215	325756
COMMWLTH OF MASS	196	704968
BONAN SEON P	188	17120
UNION SQ APARTMENTS	160	14042

Largest landowners - After 2 UPDATES of the lookup table

OWNER	PARCELS	VALUE (K)
CITY OF BOSTON	2876	3883239
CITY OF BOSTON BY FCL	1258	19117
BOSTON REDEVELOPMENT	605	164991
SANTORO VINCENT J	309	23087
BOSTON UNIVERSITY	283	300140
COMMONWEALTH OF MASS	235	528891
BOSTON HOUSING AUTHORITY	215	325756
COMMWLTH OF MASS	196	704968
BONAN SEON P	188	17120

### III. Project Goals, Objectives, and Description

The proposed collaboration will prototype and test an intelligent middleware approach that allows reinterpretation of official data and the integration of local knowledge, along the lines suggested in the illustrative parcel ownership example. The collaborators include local and regional agencies with in-house GIS expertise and significant experience acting as ‘data centers’. In particular, the Metropolitan Area Planning Council (MAPC) is the regional planning organization for metropolitan Boston, and the Boston Department of Neighborhood Development (DND) is a city-level agency focused on neighborhood development. They are ready and willing to participate in the development and testing of data-intermediary approaches that can facilitate both horizontal data sharing across agencies and towns and vertical data sharing among regional, local, and neighborhood organizations. They are interested in methods that can enhance the reusability of administrative data and streamline its integration with locally generated neighborhood data, to permit customization of neighborhood indicators and analyses.

In order to provide concrete milestones and support for realistic neighborhood planning activities, we will focus initially on datasets and middleware tools that can assist in analyzing and evaluating the economic viability and redevelopment prospects of inner city neighborhoods. The idea is to facilitate exploration—by neighborhood planners, community representatives, and business interests—of the potential for commercial, retail, and employment activity in and around their neighborhoods. The middleware will help integrate local information and reinterpret generally available public and private data about parcel ownership, land use, demographics, business and employment location, and journey-to-work. For example, one likely context is assistance in evaluating the viability and growth potential for neighborhood ‘main streets’ – that is, local commercial and retail strips within urban neighborhoods. In order to do this, we need to understand the market potential in terms of the relevant supply and demand for economic activity and opportunities for existing public investment in the neighborhood. On the demand side, we’d like to know how many persons have (or could have) ready access to the area, and what is their purchasing power, age group, etc. On the supply side, we’d like to know what types of land use and economic activity are currently in the area; what gaps exist (compared with a standard set of neighborhood ‘needs’ and activities); who controls what portions of the land; what land use and zoning changes are possible and desirable; who frequents the area daily (or at night); what worker pool has ready access; and other relevant factors

The development of a unique and elaborate model of community and economic development is beyond the scope of the proposed project – and is, in any event, not the primary focus of the effort. The main idea is to make it much easier, for both simple and complex market analyses, to tap existing administrative databases and customize their analyses and presentations so they reflect neighborhood knowledge and beliefs in ways that are defensible, efficient, and replicable. We can begin this effort by developing the tools suggested in the BRA ‘spelling’ example and testing the ease with which they can be used to encapsulate useful middleware for the target agencies and clients. Such an investigation requires both an institutional and a technical analysis, since the choice of technical sophistication and middleware components can be sensitive to the in-house skills, data sharing motivation, data maintenance strategies, inter-organizational cooperation, and participatory planning interests of the organizations involved.

Parcel data indicating land use, assessed value, and ownership is high on the ‘wish list’ of desired datasets for most neighborhood organizations. Facilitating the sharing and re-interpretation of parcel-level data is a key component of many neighborhood analyses. A growing number of assessing offices in cities, towns, and counties are providing annual, summary snapshots of their parcel data in an electronic form that is available to the public. Correcting spelling errors in these ‘official’ datasets is part of the data processing that is needed to enhance the usefulness of parcel data sharing – but it is only the beginning. Often, we’d like to categorize land in the vicinity of a proposed neighborhood development site in ways that help us determine who has what interests, what land use patterns exist and are being proposed, etc. Answering these questions involves categorizing or re-interpreting land use and ownership data in ways that differ from what is pre-recorded in the official parcel records. We might want

to identify, for example, all the city-owned land *across all its agencies*, any land in arrears on taxes, or clusters of actual (or zoned) land use that match typical patterns for neighborhood ‘main streets.’

For each of these cases, we can encode the categorization of parcels using the same ‘lookup table’ approach that handled the spelling errors. Different neighborhood organizations, or even different departments in the same agency, might well want to use a different categorization scheme reflecting their particular interests and knowledge. The proposed middleware approach makes such customization practical by encapsulating each organization’s local knowledge in the form of ‘business rules’ and lookup tables that can be documented and more easily shared, augmented, and reused. By treating this encapsulated knowledge as a plug-and-play module, rather than as permanent corrections to the official datasets, we can make dataset use much easier for data intermediaries like MAPC and DND. They can more readily accommodate parcel data updates and they can customize the analyses and data services that they provide in ways that reflect the particular needs and local knowledge of their constituents. For example, both MAPC and DND often provide communities with maps of open space, community assets, land use, and the like. In these cases, most ‘map layers’ and much of the map-production process in their data processing pipeline could remain unchanged regardless of which container of encapsulated ‘local knowledge’ is used to shade and summarize the land use and ownership patterns on the map. Having a modular approach ‘under the covers’ of the map production process would make it much easier to build a website that offers user-flexibility in choosing the method for categorizing parcels. A somewhat more sophisticated user could even upload their own lookup table and business rules for parcel interpretation.

Similarly, neighborhood market analyses of economic development opportunities, job accessibility measures, and the like could also benefit from the same strategy of building lookup tables and encapsulating classification rules. The core datasets might be different (using, for example, private sector business location databases). However, the same need remains to categorize and cluster activities (*viz.*, businesses) by type and location in a manner that allows local flexibility while being easily integrated with other map layers and analysis steps.

Considerable research and experimentation are needed to demonstrate the broad practicality of this middleware approach and to identify appropriate and sustainable strategies for decomposing the data processing steps into reusable modules and for adding ‘intelligence’ to the data intermediary. Changing from a focus on data sets to a focus on coordinated data services is especially difficult since the relevant technology is evolving rapidly and, however attractive it may be, a modular design is only as good as its weakest link. Front line agencies such as MAPC and DND are under constant pressure to address day-to-day mapping and analysis needs and have little room for experimenting on their own with new paradigms for their data processing and analysis.

It remains to be seen how far we can get in integrating and automating customizable plug-and-play components in ways that are practical and sustainable. But, with appropriate university involvement and the requested funding from the Urban Markets Initiative, the collaborating partners are ready, willing, and able to begin, and we are confident that significant steps are possible. We do not need a full suite of online, interconnected web services for the practicality of this approach to be worthwhile. The key is to treat ‘official’ data routinely obtained from other agencies as ‘read-only’ datasets, and to structure the local lookup tables and data processing steps as standardized plug-and-play components. As the shared datasets become larger and spatially more detailed, the need for streamlining the data integration and cross-referencing steps is becoming more apparent. Agencies working at the neighborhood scale and having experience in a data-center role are beginning to recognize the complexities of data coordination and maintenance and are increasingly sympathetic to the call for a new data sharing paradigm<sup>4</sup>.

---

<sup>4</sup> The semantic and ontological aspects of data sharing are active research areas in computer science. E.g., see <http://www.w3.org/2001/sw> - the ‘Semantic Web Activity’ page of the World Wide Web Consortium. As the semantic web and resource definition framework (RDF) ideas mature, they will provide standard frameworks for Web support of the types of middleware tools suggested in this proposal.

With the ‘right’ datasets in hand, we can take a good first cut at integrating the ‘local knowledge’ components into typical mapping and analysis tasks by using standard GIS tools<sup>5</sup>. It is also relatively straightforward to build a web-GIS version that would publish the statistics and maps for such analyses on the Web<sup>6</sup>. But current web-GIS tools have limited flexibility for augmenting and reinterpreting the analyses based on local knowledge. The proposed work addresses this limitation by developing a prototype system that includes:

1. ‘Lookup’ tables and processing rules for reinterpreting and categorizing (a) publicly available parcel ownership and land use data, (b) commercially available business location data, (c) socio-economic data from census and the Boston Indicators project, and (d) job classification and transportation mode choices in publicly available journey-to-work data,
2. Middleware for managing the lookup tables, uploading and editing them via the Web, and controlling their use,
3. Geospatial processing services to cross-reference, aggregate, and map the neighborhoods and the relevant service areas, business and population clusters, commute and labor sheds, etc.
4. User interfaces for four levels of use: neighborhood planners, local residents, (potential) employers, and system administrators.

Initially, the middleware will be developed on MIT servers such as the one that has been providing customizable web-access to Boston area orthophotos since the mid-Nineties. Many basemaps and data layers are already online and available on MassGIS servers via Open GIS-compatible Web Mapping Services. The ‘journey-to-work’ data are available for an ongoing study of issues of job-housing balance in greater Boston and Atlanta. Only the commercially available business-location data will have to be acquired. Whatever software is developed for the project will be freely available and, wherever possible, will use open standards and interoperable application-programming interfaces.

For the proposed intelligent middleware to be effective, sustainable, and replicable, a number of agencies and organizations should be involved in its design and in the development of the partnerships and cooperative arrangements that will be required. The plan for these institutional arrangements and practice-based research are discussed below after identification of the major partners.

#### **IV. Collaborators and Qualifications**

The primary collaborators are MIT, the Metropolitan Area Planning Council (MAPC), Boston’s Department of Neighborhood Development (DND), and The Boston Foundation (TBF). MAPC is the regional planning organization for Metro Boston and DND is a lead agency concerned with neighborhood development in Boston’s core city. We also have ongoing working relationships with MassGIS (the State’s GIS agency), the Boston Redevelopment Authority (BRA), the Metro Mayors’ Coalition (of inner city Mayors), the MassCDC Association, and several other mapping and data collection agencies and NGOs within Boston’s urban core.

MAPC and TBF are the lead partners in the Boston Indicators Project that has been accumulating and publishing socio-economic and cultural indicators for Metro Boston. MAPC has also initiated a two-year participatory planning process called MetroFuture and is in the process of designing a Regional Data Repository for greater Boston. The Metro Mayor’s Coalition and the MassCDC Association are working with MAPC on MetroFuture, and we plan to illustrate the use of our middleware tools by focusing on neighborhoods where economic redevelopment is a priority for both these two groups. Through this project, MAPC hopes to test the extent to which it can serve as an information intermediary and proactive regional

---

<sup>5</sup> Recently, an MIT Master of City Planning thesis written by Hideo Sakamoto and supervised by Prof. Ferreira experimented with such GIS models: “Socioeconomic Topography: Inner City Economic Development and Geographic Information Systems,” MIT MCP thesis, 1999, <http://web.mit.edu/uis/theses/sakamoto>.

<sup>6</sup> MassGIS, the state’s GIS agency, provides many of the basemaps used by MAPC, DND, and other local agencies. Most MassGIS datasets are available online through their web mapping services: <http://www.state.ma.us/mgis/mapping.htm>.

planning agency for its member towns and community groups without becoming burdened as a traditional data center. The Boston Foundation wants to have a more easily sustainable data processing pipeline beneath its well-regarded Boston Indicators project and wants to encourage effective data-sharing across the 101 cities and towns within the MAPC region (i.e., greater Boston).

The proposed middleware will help MAPC in managing various datasets and transforming these datasets into actionable knowledge for local planning and decision making. Parcel data is a good example of a disparate data set: MAPC has 101 cities and towns within its region and as many different assessing databases. For a planning agency, up-to-date information about land use is critical. Middleware will help MAPC in developing a coherent, region-wide layer of tax assessing data. This land-use information can also inform regional transportation models as well as help to more effectively target long-term regional policy making and public investment.

Boston's Department of Neighborhood Development (DND) is often asked to coordinate its development initiatives with other agencies that own, manage, or otherwise regulate real property in the City of Boston. Querying the city's assessing database for public or institutional owners is difficult, partly owing to the problem of variations in spelling as discussed earlier. Occupancy codes and exempt codes can help with the institutional categorization, but they are not used consistently, especially as regards exempt property. Property held by the Boston Redevelopment Authority, for example, should have an occupancy code of 903 or an exempt code of O, and other city-held property should have an occupancy code of 902 or an exempt code between A and V, but other codes may be used as well. Finding property owned by the Boston Archdiocese or any of the local universities is also fraught with the same problems. Middleware with decision rules for properly classifying these properties could make them easier to identify, assisting agencies like ours in working with other area organizations to invest public resources more strategically.

MassGIS has one of the few state GIS data repositories that supports web-feature services and web-mapping services that comply with the new interoperability standards of the Open GIS Consortium<sup>7</sup>. MassGIS is actively encouraging parcel record standardization across local communities. They have initiated outreach efforts to improve the quality and compatibility of many local datasets and they are interested in testing the practicality of the proposed middleware approach.

MIT is well qualified to play a lead role in developing the prototype information infrastructure. The MIT Urban Studies and Planning Department is world renowned. Prof. Ferreira heads the Urban Information Systems group, founded the Computer Resource Lab within MIT's School of Architecture and Planning, and has been researching urban applications of GIS since long before the advent of the Web. He is a past President of the Urban and Regional Information Systems Association and he has worked with MassGIS, MAPC, TBF, BRA and the Open GIS consortium on several GIS and database projects involving classwork, theses, and funded research. Prof. Ferreira's short resume is appended to this proposal. MIT's information infrastructure is ideal for experimenting with interoperable geospatial services. Two other MIT Planning Faculty, Prof. Lorlene Hoyt and Prof. Ceasar McDowell, are also currently working on Neighborhood Information Systems with community organizations in greater Boston. We anticipate including in this project at least one of these NGOs, most likely the Dudley Street Neighborhood Initiative (DSNI) located in the Roxbury section of Boston.

Other organizations have also expressed interest in collaborating with us and testing the 'intelligent middleware' approach. These include: other NNIP partners and participants in the Community Statistical System network (e.g., Tom Kingsley at the Urban Institute and Michael Barndt at the Nonprofit Center of Milwaukee); Federal agencies involved in data sharing and dissemination efforts (e.g., Cavan Capps at the US Census Bureau); other groups submitting proposals to UMI (e.g., Cheryl Parker at Urban Explorer); and researchers studying sustainable strategies for accumulating and retaining detailed 'city knowledge' about the built environment (Prof. Fabio Carrera at Worcester Polytechnic Institute). During the first year, through

---

<sup>7</sup> See the OpenGIS Specifications part of their Documents page: <http://www.opengis.org/specs/>

some of the activities described in the next section, we plan to identify appropriate collaborations and design our middleware tools so that they can be applied in a broad range of metropolitan settings – not just for Boston. Depending upon available funding, we anticipate expanding the initial list of partners after the first year. Many of the proposed activities during the last seven months of the project involve outreach and communication efforts to test the generality of the methods we develop and to facilitate their use by a broader community.

## **V. Milestones, Timeline, and Management Plan**

The project would begin July 1, 2004, and last 19 months through January, 2006. There would be five phases with these broad milestones:

Phase I: Summer 2004 – project planning, data sharing agreements and test services, baseline client interviews, system design, conceptual paper

Phase II: Fall 2004 – system development, advisory board meetings/demonstrations, accumulation of ‘local knowledge’, planning for the Spring phase of, project website development with discussion forum and (limited) system demonstration

Phase III: Spring 2005 – 2-day January (IAP) workshop; system testing throughout Spring via class workshop with MAPC and community clients; two or three theses on system design, ‘main street’ evaluation model, institutional analysis, and system evaluation

Phase IV: Summer 2005 – system rewrite, documentation, installation at MAPC and/or MassGIS equipment, papers and conference presentations, advisory board meeting

Phase V: Fall/Winter 2005 – outreach, evaluation follow-up, institutional recommendations, partnership expansion

During the Summer of 2004 (Phase I), we will focus on detailed planning of Phases II and III and setting up the hardware, software, accounts and data-sharing arrangements among the primary partners. The conceptual paper will explain the key ideas, system design, and illustrative examples in enough detail to facilitate Fall planning, inform our advisory board, and motivate the local agencies and community organizations who will be involved in the first year activities. We will also do the baseline interviewing of key persons in our collaborating agencies and NGOs, and confirm members of our advisory board.

During Phase II (Fall, 2004), we focus on system development and the initial round of local knowledge accumulation – that is, populating the ‘lookup tables’ and developing the processing rules that facilitate cross-referencing of data. We will also be planning for the January (Independent Activities Period) workshop and the Spring class project that will test the use of the system and tools that we develop. Finally, we will convene the advisory board in September. We anticipate having ten to twelve people on the advisory board. About half will be local – including other MIT faculty working on community development and neighborhood information systems, senior staff from the collaborating agencies (who are not directly involved in the project work), and representatives from other local agencies and NGOs who have expressed interest in the project (e.g., staff at the Boston Redevelopment Authority, MassGIS, and the Dudley Street Neighborhood Initiative). The other non-Boston members of the advisory board will be drawn from the Brookings Institution NNIP partners, Federal data-sharing efforts, the Open GIS Consortium, and the Community Statistical System network.

Phase III includes the two-day IAP (MIT's January intersession) workshop and the Spring-semester class project. The two-day workshop is an opportunity to pull together all the collaborators to examine, test, and critique the middleware tools and data services that have been developed and integrated for use during the Spring project. We also hope to have some of the outside advisory board members present during the IAP workshop. A limited amount of travel money is budgeted for this purpose and we expect that others will be able to cover their travel costs to Boston. The focus of the Spring class project effort is to test the usefulness

of the middleware tools and data services in a client-oriented setting that focuses on neighborhood development. We plan to work with at least one local neighborhood organization (that has ongoing involvement with MAPC and DND). The most likely context is to assist in evaluating the viability and growth potential for neighborhood 'main streets' and/or opportunities for CDC redevelopment of city-owned land. We will try to schedule an advisory board meeting in May to coincide with the end-of-semester class project presentations and/or the thesis defense for students whose theses relate to the project.

The Summer of 2005 (Phase IV) focuses on digesting the results of the Spring experiments and on rewriting and documenting the middleware tools and data services to reflect what we have learned to facilitate portability to other metropolitan settings.

The Fall and Winter of 2005 (Phase V) focus on outreach, further evaluation, and facilitating broader use of our results. By this time, we anticipate releasing Version 1 of our prototype tools. We also anticipate expanding our partnership and submitting additional proposals to various funding sources both (a) to help MAPC, DND, and other collaborators expand our 'data intermediary' model for the sustainable data services and (b) to contribute to a broader dialogue about scalable metropolitan information infrastructure strategies and their impact on local and regional planning.

The principal collaborators will meet monthly at MIT or MAPC, and project staff will work primarily at these two locations. We will use an MIT project mailing list for most correspondence and two other mailing lists for communication with our advisory board and with other researchers and neighborhood organizations that express interest in the progress of our effort. We will also maintain a project website at MIT. The Phases of the project and the academic semester schedule provide convenient milestones for completion of the project stages. We anticipate sending brief progress reports to the Brookings Institute at the end of each phase.

The premise underlying our project is that our 'intelligent middleware' approach is an innovative and effective alternative to the proliferation of 'data centers' and to the increasingly labor-intensive data manipulation burdening these centers. The ultimate test of this hypothesis is the extent to which this new paradigm catches on. We will not know how extensively it has caught on, of course, until long after this project is finished; however, we will be able to get some sense of the viability of our alternative by examining the use and maintainability of our prototype. We will measure how easy it is to build and maintain the lookup tables and processing rules; we will observe any difficulties involved in transferring the technical know-how to collaborating agency staff; and (via the class project) we will examine how effective the tools prove to be in facilitating useful dialogue and use of local knowledge when engaging neighborhood organizations in community development discussion. Another measure will be the interest shown in our effort by other UMI projects, NNIP partners, and other agencies that have been involved in the Community Statistical System network.

Project evaluation will include structured interviews of project partners and community clients before and after the Spring 2005 workshops. We will also keep track of how frequently the various system features are used and we will analyze the nature and extent to which local knowledge is accumulated and how it alters the perceived patterns and interpretation of land use, business clusters and gaps, job accessibility measures, etc. We will also use the system to generate market-potential indicators for the dozen or so 'main street' projects supported by Boston in the past and then compare the projects' indicated potential with their perceived success or failure. Finally, we will report the system design tradeoffs that we encounter, as well as the technology transfer experiences and collaborative planning issues that become evident.



# Metropolitan Area Planning Council

60 Temple Place, Boston, Massachusetts 02111 617/451-2770 Fax 617/482-7185

*Serving 101 cities and towns in metropolitan Boston*

April 21, 2004

Professor Joseph Ferreira  
Department of Urban Studies and Planning  
Massachusetts Institute of Technology  
77 Massachusetts Avenue, Bldg 9-532  
Cambridge, Massachusetts 02139

Dear Joe,

We are pleased to collaborate with MIT in support of your proposal to the Brookings Institution's Urban Markets Initiative to develop "Intelligent Middleware for Understanding Neighborhood Markets." The Metropolitan Area Planning Council (MAPC) has over 40 years of experience in providing data and maps to our 101 city and town membership in Greater Boston. MAPC's GIS Lab and Metro Data Center have worked with municipalities to utilize data and build statistical systems that facilitate a deeper understanding of neighborhood, municipal and regional trends to inform planning policy in the metropolitan Boston area and the impacts of those planning decisions.

MIT's proposal can assist MAPC in our mission to make the Boston region the most livable area in the nation by helping communities make strategic investment of public resources. We rely on the databases of multiple public agencies to inform our efforts. As MIT's proposal points out, these databases were not necessarily designed for integrative or analytical purposes. Developing comparable cross-town data is a crucial part of our regional planning efforts and requires a considerable amount of effort and staff time. We look forward to working with MIT to develop 'middleware' tools that might automate much of this work and help in providing information and analyses that are more readily customized to illuminate community needs and take advantage of local knowledge.

We have allocated \$19,500 in match funding for year one and \$5,000 for year two; this would accompany the \$39,080 and \$8,510 in years one and two we are requesting from the Brookings Institution.

MIT's proposed project is both timely and relevant as MAPC is currently planning ways to better provide data, technical assistance, and analysis to our 101 municipalities in light of recent technological advances. MAPC is launching an effort to build a virtual Regional Data Repository, which will be the region's premier public source for data and data services within the next five years. The tools that MIT is proposing to develop would help to lay the necessary groundwork for this important regional capacity.

Sincerely,

Marc D. Draisen  
Executive Director



INNOVATION. INFORMATION. IMPACT.  
75 Arlington Street, Boston, MA 02116 617-338-1700 fax 627-338-1606

April 26, 2004

Professor Joseph Ferreira  
Department of Urban Studies and Planning  
Massachusetts Institute of Technology  
77 Massachusetts Avenue, Bldg 9-532  
Cambridge, Massachusetts 02139

Dear Joe,

The Boston Foundation is pleased to write in support of MIT's proposal to the Brookings Institution's Urban Markets Initiative to develop "Intelligent Middleware for Understanding Neighborhood Markets."

As Director of the Boston Indicators Project at the Boston Foundation, Greater Boston's community foundation, I am all too aware of the need to develop effective "middleware" to address some of the hurdles that now impede development of fully integrated, flexible, cost-effective and sustainable data sharing efforts at the local and regional levels.

As partners with MIT and the Metropolitan Area Planning Council (MAPC) in development of a "new generation" Regional Data Repository, we have been working for a number of years to create the capacity to make relevant and up-to-date data and information available to a range of constituencies working within and across neighborhoods, municipalities and at the regional scale. We are very committed to development of the shared data resource and have already significant invested funds and staff time in its conceptualization. However, without an Urban Markets Initiative grant from the Brookings Institution to support the intense and difficult work required to overcome the technological hurdles we have encountered, we will remain in the "old paradigm" of expensive and inefficient data warehouses.

The Boston Foundation will make available a significant amount of our staff time from key personnel on from Boston Indicators Project, specifically the Project Director and Research Associate, as well as the Boston Foundation's Director of Web Communications. We will also contribute meeting space and support, and access to a number of our grantee organizations for beta testing as necessary.

I sincerely hope that the Brookings Institution will give this important project every possible consideration. Thank you very much.

Sincerely,

Charlotte Kahn, Director  
The Boston Indicators Project  
The Boston Foundation  
75 Arlington Street  
Boston, MA 02116



# DEPARTMENT OF NEIGHBORHOOD DEVELOPMENT

BOSTON'S PUBLIC FACILITIES DEPARTMENT

THOMAS M. MENINO, MAYOR

CHARLOTTE GOLAR RICHIE, CHIEF AND DIRECTOR

April 21, 2004

Joseph Ferreira, Jr.  
Professor of Urban Planning and Operations Research  
Massachusetts Institute of Technology Rm. 9-532  
77 Mass. Ave.  
Cambridge, MA 02139

Dear Professor Ferreira:

This is to confirm our support and intention to work with you in connection with your proposal to the Urban Markets Initiative of the Brookings Institution to develop computer-based tools for enhancing the accessibility and usability of public databases.

The proposed scope of work can assist us in our mission "to make Boston the most livable city in the nation by working with communities to build strong neighborhoods through the strategic investment of public resources." As part of this mission we work with other agencies that own, develop, or regulate real property in Boston's neighborhoods. We also rely on their databases to inform our efforts in developing strategic programs and policies for community revitalization. As your proposal points out, these datasets were not necessarily designed for integrative or analytical purposes, and enhancing them requires a considerable amount of effort and staff time. We look forward to working with you to develop 'middleware' tools that might automate much of this work and help in providing information and analyses that are more customized to match community needs and better able to integrate local knowledge.

Sincerely,

Larry Braman  
Manager, Mapping and Data Services

