A GIS IS BORN!
François Tessier, INRS-Urbanisation
Michel Beaudry, INRS-Urbanisation
Richard Toussaint, Ville de Montréal

ABSTRACT
This paper describes the use of SAS/BASE®, SAS/AF®, SASGRAPH® and SAS/TOOLKIT™ in creating a research oriented geographical information system for the City of Montréal. It discusses the production of a cartographic interface that combines the power of SAS/AF and SAS/GRAPH and the experience acquired by using SAS/TOOLKIT™ in providing additional geographical manipulation procedures such as «area description» and «area inclusion».

INTRODUCTION
What distinguishes a GIS¹ from other types of information systems is its spatial analysis functions. These functions use spatial and non-spatial attribute data in the GIS database to answer questions about the real world.

The original stated purpose of BDIU was to provide planners and managers of the City of Montréal with a database containing socio-economic data. The unstated purpose was to give users not only access to a wealth of data but also to provide an even easier access to methodological expertise and solutions for the manipulation of data that could resolve their research problems quickly and efficiently without the need for the user to become an expert in any other field than his own.

A SHORT STORY
The birth
BDIU was born in October 1987. This project involves the City of Montréal which is the second largest metropolitan area in Canada (3 millions inhabitants) and INRS-Urbanisation. The latter is a University of Quebec research center. The project has had the benefits in its earliest period of joining a theoretical point of view with the day to day reality of city planners in dire need of data to base their decisions.

The childhood
It is during this period that modelization then data collection, acquisition and integration took place. While all of those tasks were being performed, the SAS® system was selected as the software environment for BDIU. This choice was based on five criterias:

- Quality and diversity of software tools
  Most of our energy was devoted to understanding and handling data, not writing code. Therefore we were looking for the most complete and reliable environment available.

- Ability to handle large volumes of data
  Most of the datasets contained over 50,000 observations and an average of 60 variables. Furthermore, the data model implemented implied a frequent use of sort and merge routines.

- Ability to build customized applications
  Both the presence of users with specialized needs and the necessity to perform certain tasks repeatedly required the capability to create turnkey applications.

- Availability on various hardware platforms
  An open approach in terms of computer platform was essential to allow for probable future system enhancements.

- Support of foreign languages
  Considering that the system needed to operate in a French environment, it was vital that any output, graphical or not, could be produced using that language.

Since the project’s origin, the role of BDIU had progressively evolved. The idea of a simple socio-economic data bank was not sufficient anymore. It needed to be transformed into the more ambitious concept of an urban information system.

Adolescence
The transition to the adolescence period occurred almost naturally as the logical outcome of the first phase of the development (i.e. the data bank). BDIU evolved to become a complete information system with:

- Documentation
- Analysis tools
- Technical support
- User training
- Methodological expertise

and eventually application development.

In order to keep up with the diversification of the requirements, software development soon became an important part of BDIU’s activities.

An example of such developments is CARTOGRAF, a menu driven application built using SAS/AF® and SAS/GRAPH® to help non specialists design and produce thematic maps.
The creation of CARTOGRAF accomplishes three objectives:

- **Fulfill users' needs**
  Cartographic representation of information was, from the beginning, one of the most demanded functionalities.

- **Provide users with a powerful instrument of analysis**
  We believed that maps should not only be an interesting way of presenting results but also a powerful tool to perform analysis.

- **Increase the system's ease-of-use**
  Most of our users make an intensive but sporadic use of the system. A user-friendly interface is well adapted to this type of access.

The functionalities provided by CARTOGRAF are divided into five categories:

1) **Definition and computation of analysis classes.**

   ![Definition of Classes](image)

   **DEFINITION OF CLASSES**
   (Attribute File: ______)

<table>
<thead>
<tr>
<th>Minimum Value</th>
<th>Maximum Value</th>
<th>Character Value</th>
<th>Labels (description)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   [ESC] CANCEL  [F10] SAVE

2) **Definition of SAS/GRAHP graphics elements such as titles, notes and legend in terms of size, color, shape, font and position on the map.**

   ![CARTOGRAPHIC REPRESENTATION](image)

   **CARTOGRAPHIC REPRESENTATION**
   Name of coordinate file: [ ]
   Name of data file: [ ]

   EXECUTE [F10]  QUIT [ESC]

   Enter the parameters then press [RETURN]


3) **Definition and positioning of other cartographic objects such as the arrow indicating north and the map scale.**

4) **Storage and retrieval of output parameters and map features using a permanent attribute file.**

5) **Integration and computation of default values in an attempt to provide users with on-line expert advice (as given by a professional cartographer) on such elements as context sensitive map attribute positioning (under development).**

Calling CARTOGRAF a dynamic application refers to its ever-improving capability accomplished through a continuous process of user evaluation. Our objective is to provide users with as much expertise as possible in their areas of interest.

We believe that CARTOGRAF and version 6.07 of the SAS system, with the introduction of the graphic editor, will enhance considerably the potential of BDIU regarding thematic map production.

**COMING of age**

BDIU's data bank contains information whose common denominator is a spatial reference. This spatial reference links textual information from the main database to other compatible geographic bases used by several departments within the City of Montréal and some research centers. The presence of such geographical information offers enhanced possibilities in terms of analysis. For example, a team of planners could define territorial divisions and question the system about them. Questions such as "What is the population density in one district?" or "What is the geographical distribution of office space in the downtown area?".

In order to help answer such questions, we decided to create two new SAS procedures using SAS/TOOLKIT™. The choice to invest in the SAS/TOOLKIT approach was based on the following reasons:

- It allowed us to maintain a uniform environment for the user thereby justifying our investment in the SAS System.
- It was less costly to adapt an existing environment than to create additional functionalities using a new system.
By selecting SAS/TOOLKIT as the software development environment, we were benefiting from the SAS Institute vast experience in building portable and maintainable code.

The first procedure developed using SAS/TOOLKIT was PROC GINFO. The procedure computes, after reading a map dataset, the minimum and maximum in X and Y, the number of zones, the number of polygons and the total number of coordinates. It also calculates for each zone the minimum and maximum in X and Y, the coordinates of the geometric centroid, the area and the number of polygons. The results are displayed in the output window and written to a dataset which can be used for further computations.

For example, the population density represented on the figure shown below was obtained using PROC GINFO's computed areas.

**SYNTAX:**

```sas
PROC GINFO <DATA=SAS-data-set>
  OUT=SAS-data-set
  <NOPRINT>

where:

- **DATA** is the input map dataset.
- **OUT** is the output dataset where the information for each zone will be written.
- **NOPRINT** is an option that overrides the production in the output window of the information for each zone.
```

**Example 1:**

```sas
PROC GINFO DATA=MTL.EAP86A10
  OUT=RESULTS;
RUN;
PROC CONTENTS DATA=RESULTS;
RUN;
```

The resulting output for this example can be found in the appendix.

PROC GINFO was the second procedure developed using SAS/TOOLKIT. Its purpose is to analyse a dataset containing geographical locations of objects such as daycare centers or dwelling units to see how they compare with the polygons defined in a map dataset. In other words, it reads \((x,y)\) coordinates from the first dataset then lists the polygons in which they are included.

**SYNTAX:**

```sas
PROC GINFO <DATA=SAS-data-set>
  MAP=SAS-data-set
  OUT=SAS-data-set
  <NOPRINT>

  <ID_ZONE zone;>
  <ID_POINT point;>

where:

- **DATA** is the input dataset containing the name of each location.
- **MAP** is the map dataset containing the description of the area.
- **OUT** is the output dataset where the information for each match between a location and a zone will be written.
- **ID_ZONE** is the name of the variable containing the zone identifier in the map dataset.
- **ID_POINT** is the name of the variable containing the location identifier in the first dataset.
```

PROC GINFO also writes the results in a dataset. The following example shows how a dataset containing the location of daycare centers in Montreal is compared with a map dataset to produce the geographical map of the distribution of daycare centers across the city.

**Example 2:**

```sas
PROC GINFO DATA=MTL.DAYCARE
  MAP=MTL.AM89MTL
  OUT=RESULT1;
  ID_ZONE ZONE;
  ID_POINT POINT;
RUN;
PROC FREQ DATA=RESULT1;
  TABLES ZONE /OUT=RESULT2 NOPRINT;
RUN;
PROC GMAP DATA=RESULT2
  MAP=MTL.AM89MTL
  ANNOTATE=ANNO
  ID ZONE;
  CHORO _COUNT_/DISCRETE
  LEGEND=LEGEND1
  FORMAT_COUNT C.;
```
Distribution of Daycare Centers by Planning Districts
City of Montréal 1992

SAS/TOOLKIT provides us with the opportunity to add user defined functionalities to BDIU’s environment and to improve our understanding of the SAS system. Unfortunately, the quality of the documentation provided with the experimental release lacked in clarity and completeness. But the patience and determination of SAS technical support helped us overcome those difficulties.

Both procedures are currently being tested in order to improve the precision of the computed results in some “pathological” cases. They were written using the C language in order to facilitate the pointer/address "gymnastics" and ensure future migration of the code. We also took great care in following the portability guidelines and not to use system dependent shortcuts.

CONCLUSION

"There are as many definitions... of GIS... as there are disciplines involved in using geographic information systems" (Huxhold and Burrough).

We believe that the three functions we have written are the necessary first step in establishing a real-world GIS. They provide our users with a toolbox enabling them to get a better understanding of the region they have to manage.

The response to their introduction was so enthusiastic that we are already planning the next procedures, more specifically one to convert non standardized postal addresses to a (x,y) coordinate system. The SAS system has proven to be a valuable tool at all phases of our project: from building and managing a database to the current evolution toward a complete socio-economic research oriented GIS.

Note

The graphical outputs accompanying this document are translated examples generated specifically for presentation purposes. Originals exist in French and are available upon request.

Acknowledgment

A special thanks to Mark Moorman from SAS technical support for his patience and his sound advice.

REFERENCES

(1) Geographical Information System
(2) BDIU is the french acronym for Banque de données et d’information urbaine.
(3) Information referenced to a specific geographical location.
(4) A dataset used by proc GMAP

BIBLIOGRAPHY


SAS®, SAS/AF®, SAS/GRAPH® and SAS/TOOLKIT® are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

CONTACT

François Tessier
INRS-Urbanisation
3465 Durocher
Montréal, Québec
H2X 2C6
Canada

EMAIL: TESSIER@INRS-URB.UQUEBEC.CA
APPENDIX

Characteristics for each zone

<table>
<thead>
<tr>
<th>Zone ID</th>
<th>Nbr.</th>
<th>Min X</th>
<th>Max X</th>
<th>Min Y</th>
<th>Max Y</th>
<th>Centroid X</th>
<th>Centroid Y</th>
<th>Area</th>
<th>Nbr. polygons</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM89X01</td>
<td>809</td>
<td>284.36</td>
<td>295.25</td>
<td>40.53</td>
<td>49.40</td>
<td>292.61</td>
<td>49.27</td>
<td>24.20</td>
<td>3</td>
</tr>
<tr>
<td>AM89X02</td>
<td>55</td>
<td>293.16</td>
<td>298.02</td>
<td>42.40</td>
<td>49.29</td>
<td>295.34</td>
<td>45.99</td>
<td>16.48</td>
<td>1</td>
</tr>
<tr>
<td>AM89X03</td>
<td>56</td>
<td>295.26</td>
<td>300.95</td>
<td>42.65</td>
<td>48.90</td>
<td>298.48</td>
<td>45.74</td>
<td>15.93</td>
<td>1</td>
</tr>
<tr>
<td>AM89X04</td>
<td>164</td>
<td>299.33</td>
<td>304.48</td>
<td>43.30</td>
<td>52.68</td>
<td>302.09</td>
<td>48.52</td>
<td>25.41</td>
<td>1</td>
</tr>
<tr>
<td>AM89X05</td>
<td>48</td>
<td>295.97</td>
<td>301.54</td>
<td>41.08</td>
<td>44.43</td>
<td>299.09</td>
<td>42.82</td>
<td>11.37</td>
<td>1</td>
</tr>
<tr>
<td>AM89X06</td>
<td>628</td>
<td>296.43</td>
<td>303.36</td>
<td>38.34</td>
<td>43.31</td>
<td>302.82</td>
<td>40.03</td>
<td>14.31</td>
<td>5</td>
</tr>
<tr>
<td>AM89X07</td>
<td>139</td>
<td>290.89</td>
<td>297.33</td>
<td>34.33</td>
<td>41.63</td>
<td>294.44</td>
<td>38.09</td>
<td>21.70</td>
<td>1</td>
</tr>
<tr>
<td>AM89X08</td>
<td>207</td>
<td>294.19</td>
<td>301.86</td>
<td>32.85</td>
<td>39.01</td>
<td>298.32</td>
<td>36.31</td>
<td>15.48</td>
<td>1</td>
</tr>
<tr>
<td>AM89X09</td>
<td>699</td>
<td>295.34</td>
<td>306.63</td>
<td>52.43</td>
<td>62.43</td>
<td>296.57</td>
<td>55.66</td>
<td>42.46</td>
<td>4</td>
</tr>
</tbody>
</table>

NOTE: After reading dataset MTL.AM89MTL:
- The total number of coordinates: 2807
- The total number of zones: 9
- Maximums x and y: 306.63 and 62.43
- Minimums x and y: 284.36 and 32.84

The SAS System 13:11 Wednesday, March 18, 1992

CONTENTS

<table>
<thead>
<tr>
<th>Procedure: CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set Name: WORK.RESULT1</td>
</tr>
<tr>
<td>Member Type: DATA</td>
</tr>
<tr>
<td>Engine: V506</td>
</tr>
<tr>
<td>Created: 13:14 Wednesday, March 18, 1992</td>
</tr>
<tr>
<td>Last Modified: 13:15 Wednesday, March 18, 1992</td>
</tr>
<tr>
<td>Data Set Type:</td>
</tr>
<tr>
<td>Label:</td>
</tr>
</tbody>
</table>

Data Set Page Size: 8192
Number of Data Set Pages: 1
Filename: SPHINX$DUB3$SASWORK.SAS$WORK210001E5$RESULT1.SASEB$DATA
Disk Blocks Allocated: 18

Alphabetic List of Variables and Attributes

<table>
<thead>
<tr>
<th>#</th>
<th>Variable</th>
<th>Type</th>
<th>Len</th>
<th>Pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>AIRE</td>
<td>Num</td>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td>6</td>
<td>CENT_X</td>
<td>Num</td>
<td>8</td>
<td>42</td>
</tr>
<tr>
<td>7</td>
<td>CENT_Y</td>
<td>Num</td>
<td>8</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>MAX_X</td>
<td>Num</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>MAX_Y</td>
<td>Num</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>MIN_X</td>
<td>Num</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>MIN_Y</td>
<td>Num</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>NBR_POLY</td>
<td>Num</td>
<td>8</td>
<td>74</td>
</tr>
<tr>
<td>9</td>
<td>NBR_PTS</td>
<td>Num</td>
<td>8</td>
<td>66</td>
</tr>
<tr>
<td>1</td>
<td>ZONE</td>
<td>Char</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>