Spatial-it practical challenges (3)

A data quality impulse

Introduction

In previous editions of GeoInformatics we've introduced the Spatial-it concept, discussed its business opportunities and looked at some practical aspects of this new approach to Geo-IT. As we've explained Spatial-it can be described as the mechanism that allows all employees in the Enterprise to benefit from the powerful visual aspect that spatial information, integrated in the everyday information systems, represents. Information presented in a geographic way, either complementing or maybe even replacing the alphanumeric data, results in faster comprehension of the information. This results in an improved efficiency of the operations and an optimisation of the decision support, and thus in measurable economical benefits.

In the previous articles we've looked at some practical aspects that introducing this technology brings about. In the December '99 issue you can read about the changing business for software vendors of CAD and GIS technologies. It also explains how the typical computer of the "new" users of spatial information and their peripherals urges the developers op Spatial-it applications to be creative and selective in presenting the spatial-data alongside the alphanumeric data on everyone's desk-top. The March '00 article looked at the various users using en benefiting from Spatial-it applications and some implementation strategies to ensure the organisation does profit form the promised benefits. In this third article on practical aspects we will start to look at one of the organisations most valuable assets, its ***data***.

Data as a prerequisite

As all ICT-professionals among you know from one of your first Informatics classes, in ICT we distinguish data from information. Data as the raw form of bits and bytes, 0 and 1's , a to z, and putting it "spatially" points, lines and polygons. This data becomes information once it's put in a context. That way the employee recognises the presented numbers as the monthly salary, or the red polygon on the screen as a representation of a polluted parcel of land.
This data is a prerequisite in every Spatial-it project. As those projects have everything to do with presenting the right information, this via the manipulation of various datasources, combining them with typical spatial-data as maps. The quality of the final result is 100% dependent on the underlying data, both spatial and alphanumeric.
Based on practical experience in our Spatial-it projects we'll look at the following aspects of the data:

- ✓ Data Quality
- ✓ Relation spatial and alphanumeric data
- ✓ Data Responsibility
- ✓ Data Aggregation

Data quality

As Spatial-it potentially allows all employees to access the available information, data-quality becomes (if possible) even more important than it ever was before. As long as data is restricted to a relatively small group of insiders, the lack of quality of data can be compensated by the knowledge of its users. They'll know some of the data isn't up-to-date, or that part of it is missing because of this or that reason. The users still know how to interpret the information presented to them, or know who to ask what questions to ensure the decision is based on "solid ground". How different the situation becomes when virtually any colleague will have access to the information and use it in unforeseeable situations. They will assume the presented information is correct, up-to-date and complete.

Thread or a challenge?

So we need to be very cautious what data we incorporate in our Spatial-it environment, but more important we need to ensure this data satisfies our quality-standards so we present useful information.

This may be felt as a *thread* by some. A reaction you might recognise by some of your colleagues is "yes, please allow me to see all the information I can lay my hand on……but why would anybody want to see my information, they wouldn't understand it, or know what to do with it…." .

But in fact it's a big *chance* for your organisation. Ensuring data quality starts at its source, the place where it enters the organisation or is created combining other data sources in your work-flows. Often you will see it takes just a little more effort to increase the data-quality at that moment. Thereto ensure that the data is complete and correct, just performing that extra check or add that bit of additional information required further down the line. In this process you'll discover that knowing that "their" information will be used corporate-wide and every omission will be revealed at large, does stimulate employees to increase the data-quality level to meet the higher standards. So there's your chance for a quality impulse, initiated by your Spatial-it project.

Relation spatial and alphanumeric data

Jet another quality impulse is initiated through the various information-relations a typical Spatial-it application encompasses. Using a Spatial basis many different data-sources will be combined, using a common key, as for instance the address, to access the underlying databases. Combining this information often isn't as forthright as we would like. Years of island-automation, focused on optimising the individual business-functions, forgetting to take the overall processes of the organisation into account, have led to stacks of different databases. These database do contain overlapping information, but on a data-level (to fall back on our "lesson one" again) this information is often represented differently. The address "Koningin Beatrixstraat" can also be stored as "Kon. Beatrixstraat" or "Kon. Beatrixstr." all referring to the same street but difficult to "join" in automated processes.

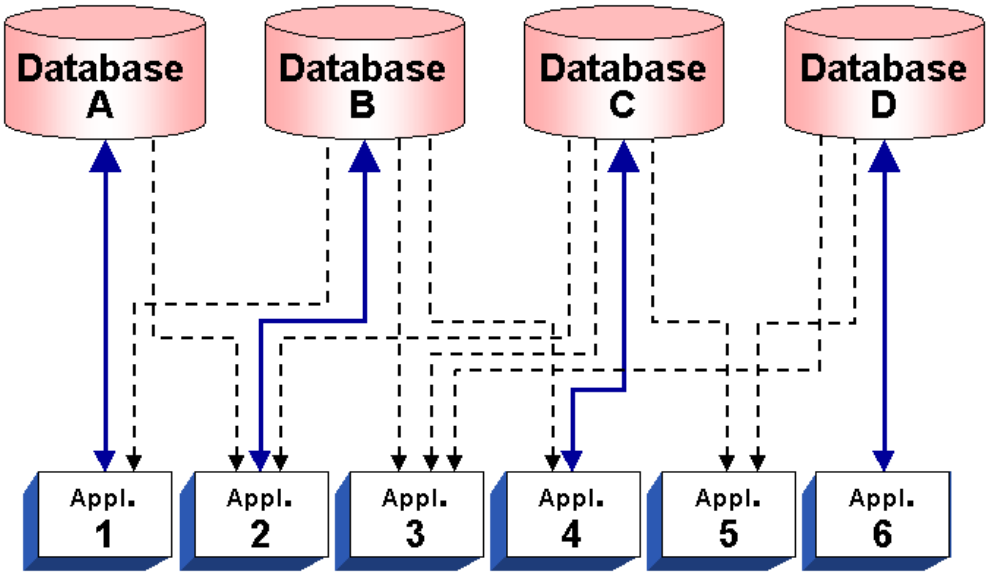
Our Spatial-it projects more than once got us involved in optimising the available data of an organisations. Particularly for the widely used data of a Spatial environment (addresses, subjects, assets) a Spatial-it project is a definite stimulant to restructure and centralise this datastorage. This results in a (virtually) central database environment which we call the "base-registration". From that bases we try to modify all relevant databases to use this central data rather than the own "island-database". Apart from data quality gains as up-to-date, complete and correct information, this also results in efficiency benefits as no longer an address is stored an maintained in say 5 different environments.

Once this database structure is in place organisations implementing additional new applications need to take this structured database into account. To fit into a corporate information structure

new off-the-shelf application need to have an open structure allowing integration and sharing of central databases. Applications build in-house will be based on this datastructure.

Data responsibility

To enable this process of “public databases”, a supporting data-management organisation becomes vital. Data-managers and data-administrators will be involved in Spatial-it projects as we want to address all or at least a fair portion of the available data in an organisation. Following the previous argumentation on data quality, the interests of data-managers en Spatial-it specialists are the same: ensuring a high data quality standard. As neither the data-managers nor the Spatial-it-specialist ultimately fill the databases, we need to work on creating the right conditions for the organisation to ensure the data is captured correctly. One of the main issues in this context is clearly define the responsibility for the various (virtual) databases. A practical approach to the question who’s responsible for what data, is illustrated in figure 1.



The figure shows a number of applications using various databases. The arrows indicate the relation between the application and the databases. A second look at the figure will show you that the we’ve used two kinds of arrows. This is where the principle of this approach becomes apparent. The definition is that every database (or virtually defined set of data-objects) is “owned“ by exactly one application (indicated by the solid doubleheaded arrow). This is the application that will capture and maintain the data. All other applications will be allowed to query the database but not to modify the information (dotted lines). The second definition in this approach is that an application can be linked to an organisational unit and via that way the responsibility for that data can be linked to this unit. In case of a reorganisation this approach, in general, allows you to easily “remap” the responsibilities.
Your spatial data fits in this schema, it can be regarded as just an other set of data and will follow the normal procedures. Often the organisational unit traditionally responsible for the CAD and/or GIS environment will be responsible for the spatial-data quality, using their spatial-data-management applications.

Data aggregation

Another interesting aspect of our “data-study” is the aggregation-level of the data. Particularly when our aim is to provide every employee with the right information to do their job, it’s legitimate to wonder what kind of data does support that goal. Simply presenting the data as it is used by the specialists in all its details isn’t always enough. On the contrary, this can lead to

misinterpretation of the data, which in fact leads to wrong information. This again would confirm the idea of some of the specialists that only they know how to read “their” data.

Let’s look at information from an environmental information system as an example. The information in such a system contains a huge amount of technical detailed data, requiring a specialist to correctly interpret this information. However if we aggregate this data to a higher level, and present it via a Spatial-it environment we see a different picture. In this context think of a presentation as in figure 2, where for the investigated parcels the system uses a couple of colours to inform the user of the essential information, e.g. Green = checked, no pollution Orange = checked, lightly polluted; Red = checked, heavily polluted; Yellow = still to be investigated. All of a sudden the information becomes easy to read and interpretable for every “layman”. And for 80% of the requests for information outside ones own profession this level of aggregation often is just right.

If we look at the information a typical managers needs, again aggregation is an indispensable mechanism to present information of a broad area in a compact comprehensible way. Although Spatial-it isn’t the only way to support managers with information to monitor processes or as decision support, you will appreciate that it does provide the ease of access, and the strength to present lots of information in just this one picture.

So aggregation plays an important role, we need to pay serious attention to the (possibly various) levels on which we need to aggregate the data in order to present the right information to the organisation.

More data …

In the next article we’ll continue this focus on different aspects related to the fundament of Spatial-it, and in fact any, application, the data…. Data storage, data repositories, data warehousing, we’ve got plenty of buzz-words to continue with.

Conclusion

Once again I’d like to conclude this article with some reassuring words. Although creating a solid “data” bases for your Spatial-it applications does require your full attention and will take up a fair bit of your time, the result will be most rewarding. So don’t be deterred. All applications of an organisation will benefit from the achieved quality impulses initiated from Spatial-it. In sharing

some of our experiences in implementing this technology with you, we hopefully offer you some handles to tackle your specific situation.

In addition to the advise gathered from magazines like GeoInformatics you may want to consult ICT-specialist to benefit from their experiences gained in other projects. Experience both on organisational aspects as on technology issues. Obviously you can find professional partner organisations that are more than willing to help you implementing this new technology. In that process be selective. As Spatial-it fills the traditional gap between GIS solutions and the other business processes it's not obvious that every GIS-solution provider can do the job. Knowing what a GIS can do is not enough. As this article explains a good knowledge of managing the mentioned data aspects is an other expertise you need to look for. So you'll be looking for system integrators with the right vision, a broad ICT knowledge and a solid background in Spatial technology, in fact a ***"****spatial integrator*" as I like to profile our company, Vicrea Solutions.

Jan Roodzand at the time of this article was director of Vicrea Solutions

2012 addition:
Jan Roodzand is now management consultant at Roodzand Advice
And managing partner at Ruimteschepper

ROODZAND ADVICE

He can be reached at j.roodzand@roodzandadvice.nl,
Tel. +31 6 200 133 65