

DATA DRIVEN SYSTEMS AND SYSTEM DRIVEN DATA: THE STORY OF THE FLANDERS HERITAGE INVENTORY (1995-2015)

Koen Van Daele, Leen Meganck, Sophie Mortier

Flanders Heritage, Brussels, Belgium - koen.vandaele@rwo.vlaanderen.be
Flanders Heritage, Brussels, Belgium - leen.meganck@rwo.vlaanderen.be
Flanders Heritage, Brussels, Belgium - sophie.mortier@rwo.vlaanderen.be

KEY WORDS: inventories, databases, cultural heritage, quality control, open source, metadata

ABSTRACT:

Over the past 20 years, heritage inventories in Flanders (Belgium) have evolved from printed books to digital inventories. It is obvious that a system that publishes a digital inventory needs to adapt to the user requirements. But, after years of working with a digital inventory system, it has become apparent that not only has the system been developed to the users needs, but also that user practice and the resulting data have been shaped by the system.

Thinking about domain models and thesauri influenced our thinking about our methodology of surveying. Seeing our data projected on a common basemap led us to realise how intertwined and interdependent different types of heritage can be. The need for structured metadata has impressed upon us the need for good quality data, guaranteed by data entry standards, validation tools, and a strict editing workflow. Just as the researchers have transitioned from seeing their respective inventories as being significantly different to actually seeing the similarities between them, the information specialists have come to the realisation that there are synergies that can be achieved with other systems, both within and outside of our organisation. Deploying our inventories on the web has also changed how we communicate with the general public. Newer channels such as email and social media have enabled a more interactive way of communicating.

But throughout the years, one constant has remained. While we do not expect the systems to live on, we do want the data in them to be available to future generations.

"To improve is to change, so to be perfect is to have changed often".¹ In Flanders (Belgium), heritage inventories have a long history. Over time, the way these inventories have been constructed has changed significantly. Once purely an analogue affair, nowadays computer systems and databases are omnipresent. Constructing a heritage inventory without a database has become unfathomable. It is obvious that a digital system to record and present heritage data needs to fit the users' requirements. But what might be less clear is that not only do the users form the system, but the system also forms the users. This paper will examine how the Flanders Heritage inventory system was shaped by the data, and how the data, and the way they were collected, was influenced by the system.

1. THE FLANDERS HERITAGE INVENTORY: AN INTRODUCTION

The surveying of architectural heritage by the government dates back to the foundation of the Belgian state in 1830. However, it was not until the late 1960s that a systematic survey of architectural heritage for the whole Belgian territory was deployed. Very soon, this project was divided between Flanders and Wallonia, as cultural affairs were considered regional. The results of the survey were published in books, *Building through the ages in Flanders* (Bouwen door de Eeuwen heen in Vlaanderen) (Hooft and Verwinnen, 2008). This regional survey for Flanders was completed in 2011, so it spans a period of almost 50 years!

In the mid-nineties, one of the precursors of Flanders Heritage, Monuments and Landscapes (Monumenten en Landschappen), started work on an online version of these books. The initial idea

¹W. Churchill, 1925, as quoted in: Richard M. Langworth, Churchills Wit. The definitive Collection, Ebury Press, 2009, p. 27.

called for a website that was an identical twin to the books. The books were diligently scanned and processed with optical character recognition (OCR) software. The layout of the books was used as the guiding principal to decide what became a separate item or page in the website. Of course, this conversion provided for very little structured metadata. For most records, all that was present was the name of the province and the municipality the item was located in. Almost all searches had to be done full-text. While this was a huge improvement upon the books, it did mean that a lot of questions remained unanswerable.

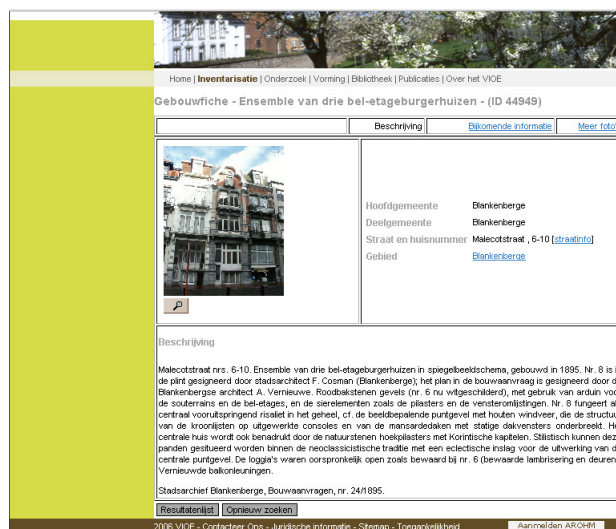


Figure 1: The inventory of architectural heritage circa 2006.

In 2004, an organisational shift took place within the Flemish

Government. Two previously separate entities, the Institute for Archaeological Heritage (Instituut voor het Archeologisch Patrimonium) and the Monuments and Landscapes Agency (Administratie Monumenten en Landschappen) were merged. The first entity was focused on archaeology, the second on architectural history and landscapes. Not only were their areas of expertise different, but they also had widely differing database systems. The archaeologists had only shortly before started developing an inventory of archaeological sites and findspots (CAI or Central Archaeological Inventory) (Van Daele, 2004). Since this was a digital born inventory, it focused on strong, structured metadata by using controlled vocabularies and a relational database. Because most archaeological sites are not locatable via address, the CAI incorporated geographic information systems (GIS) from the beginning.

This led to a new version of the inventory of architectural heritage, launched in 2009. This version mixed the older written descriptions with newer spatial data and keywords from the thesauri. It was to be the first dataset in a heritage portal. The premise was that every discipline or sub-discipline had very different wants and needs, but that they all were heritage and should be presented to the public together. Over the years, datasets for musical organs, World War I relics, historical parks and gardens, heritage trees and shrubbery, heritage ships, etc. were added. While this heritage portal has been successful as a website, it has also fundamentally altered the way that heritage professionals view their data and even their heritage.



Figure 2: The inventory of historical parks and gardens, 2015.

2. FINDING TREASURE: FROM FULL-TEXT-SEARCH TO THESAURI AND STRUCTURED LOCATION DATA

When the first database went online, the possibilities for search increased dramatically. Instead of having to dig through stacks of books to find information, researchers could now query the database. Since this database was very much a digital version of the paper books, the search interface was very much oriented towards full-text search.

While this was in itself a huge improvement, the need for structured metadata became apparent. This led to a greater understanding and appreciation of the methodologies involved. For the database of architectural heritage, two main needs were detected.

First and most importantly, a strong need was felt for up to date information on the location of the architectural heritage. Up to that point, all that was available were textual strings describing

the address or the general location of the heritage. (Van Lindt et al., 2006) A large-scale project was undertaken to generate structured location data. This involved matching the textual data to the Central Addresses Reference Database (Centraal Referentie Adresen Bestand or CRAB) when possible and georeferencing based on these addresses. Of course, for a significant amount of records, manual corrections had to be undertaken. This happened with addresses that no longer existed, or with structures that had no address to begin with (e.g. because they were situated on public domain). Generally, the older the book the information originated from, the more manual work that was needed. By 2009, the entire inventory of architectural data had been provided with spatial data. (Hooft, 2011)

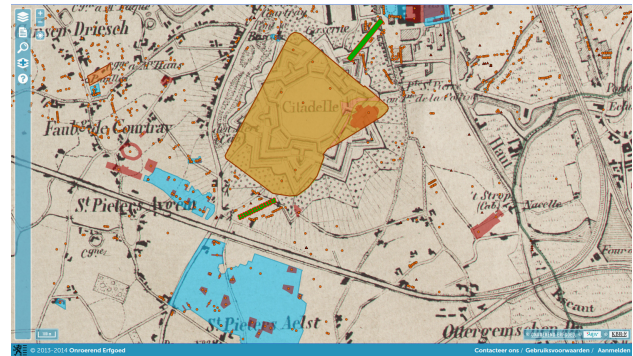


Figure 3: Different types of heritage in Gent as projected on the Vandermaelen map (1850).

At first, the portal used this data sparingly. Individual items were illustrated by a small interactive map, and a nightly export of the spatial data to a Shapefile or a KML file was offered. As time progressed, the tools for presenting GIS data on the web matured and a geospatial portal (<https://geo.onroenderfgoed.be>) was added. Now, this portal serves as an excellent way of presenting the different inventories together. They can be viewed alone or together, with modern day or historic basemaps, and in combination with other useful spatial data such as zoning plans. It is geared towards spatial planners, actuaries, and local communities, and provides a very straightforward look at heritage in Flanders. It also makes it much easier for a heritage researcher from one discipline to survey the surroundings he or she is working in and to discover what other heritage could influence his or her own area of work. This reinforced the growing idea that maybe the different types of heritage were not so different after all.

Secondly, the government wanted to shift from a geographical approach to a more thematic approach with inventories and surveys. While previously an entire community would be screened, it was determined that it should be possible to do a survey of a certain theme, e.g. religious architecture. But the current dataset made this very difficult since a query for church” would also return records about the Churchstreet”, but it might leave out a record about a cathedral”. To counteract this, it was decided to adopt a controlled vocabulary for the database, much as the archaeologists were doing in the CAI. While the CAI made use of simple controlled lists, a more comprehensive approach by using thesauri was chosen for the architectural heritage.

This led to a gigantic undertaking. All 70.000 records in the database had to be screened and assigned terms from the vocabularies to designate what type(s) of building each record represented, from what period it originated and in what style it was built. The actual thesauri themselves were drawn up over the course of this project by combining original work with thesauri already

available such as the Art and Architecture Thesaurus (Getty Research Institute) and the Thesaurus of Monument Types (English Heritage, now Historic England).

Apart from the fact that all these heritage records were now searchable by type, date, and style, constructing the thesauri themselves also caused a significant rise in understanding how we all viewed our heritage. The thesauri were designed by committee. Fine-tuning labels and scope notes for certain concepts led to heated debates and made it clear that researchers sometimes held widely differing beliefs on the meaning of certain labels and were wholly unaware of this fact.

While the first three thesauri were designed with the architectural heritage in mind, the idea was quickly adopted by the other datasets that were added to the heritage portal. Interestingly enough, at first it was thought that there should be both an architectural and an archaeological thesaurus. But quickly it was realised that a church is a church, no matter if it is still standing or only traces of it remain buried under a parking lot. Expanding the thesaurus of building types with types of archaeological sites again led to interesting discussions and an extended mutual understanding. Every new dataset that was added to the portal caused a similar pattern of knowledge acquisition.

Visualising the different types of heritage together and consolidating the controlled vocabularies used into one overarching heritage thesaurus helped to make it clear to the heritage researchers that the different types of heritage might not be so different after all.

3. FROM INTEGRATED DATA TO INTEGRATED SURVEYING

As more inventories were added to the heritage portal, it became clear that they could share much more than just a presentation layer. While different inventories sometimes referred to the same information with different labels, they all captured the same types of attributes: What? Where? When? Why? Instead of just integrating the presentation of the different databases, the actual data management tools and interface were being integrated. At first this led to different databases all having the same data model, but different content. Progressively it became clear that we could not only use the same data models, but maybe also the same database.

Because every database used the same data model and the same thesauri it became possible to launch queries across the datasets. It was now possible to find all manors, independent of the dataset they were a part of. While this was again a very powerful feature, it also pointed towards a creeping problem. In the past all inventories and surveys were being run more or less independently and were meant to be consulted independently as well. Naturally this approach led to a certain amount of duplication. The inventory of architectural heritage would describe an 18th century manor house and include a small amount of information about the gardens surrounding the house. The inventory of parks and gardens would describe the gardens at length and include a short summary about the manor house. And in the archaeological inventory, this same manor house might be present because it contains the vestiges of a medieval castle. Now, the user who looks for this manor house might find three records, all containing similar but not quite identical information. This was a major indication that we should go further than just integrating the data models. We also needed to integrate the data itself.

This new insight led to a new approach on data management for which we are currently migrating the each dataset towards one integrated heritage database. This database gathers information

on all types of heritage without discriminating between different disciplines. The need to label certain objects as landscape or archaeology becomes more and more superfluous once you start looking at it from different viewpoints. A building can easily have an archaeological component.

The basic idea is to remove all partitioning walls between the different datasets. This new integrated database gives Flanders Heritage the opportunity to start new inventories and research from a truly interdisciplinary point of view. Heritage landscapes are no longer seen without their archaeological context and vice versa. Buildings and structures are no longer seen as separate entities, but also as formative elements in the surrounding landscape.

The actual surveying is no longer limited to a certain discipline, but is carried out by specialists from different fields. A manor house will be surveyed by a team of architectural heritage specialists and experts in historic gardens or trees, while an archaeologist will screen the site for important archaeological elements. Similarly, when landscapes are being researched, the archaeological potential will be assessed and the farms that helped form the cultural landscape will be examined for their architectural value. While this interdisciplinary way of working creates many interesting synergies, it also requires a lot of planning and communication since the calendars of all concerned parties need to be synchronised.

4. QUALITY CONTROL IN A DIGITAL ENVIRONMENT

Anyone who has ever worked with a database knows how important data quality is. This is commonly known as "*garbage in, garbage out*". Only when the data in the system is of good quality will the results produced by querying and searching make sense. Several factors play into this process. Not only does the data that's being entered into the system need to be logical and consistent, it also needs to take into account what the desired output is.

Maintaining a live database on the web also produces its own set of problems. As long as the inventories were being maintained in print they were as good as immutable. All parties involved knew that a certain book was published in a certain year and could understand that the information contained in it might have become less correct since then. But, with a database or website, it is expected that the information is accurate and up to date. This can be as simple as marking a building as demolished or changing the address it is located at. Or it can be a matter of indicating how old a certain aspect of the information record is. Keeping more than 100.000 records up to date like that is a daunting task in itself.

In all of this, good communication is quintessential. Not only do the data registrars need to communicate about what they enter and how they enter the data, the heritage professionals and information experts need to be in constant contact to ensure that the entry of data will lead to the desired results in the long term.

With a large group of researchers responsible for data input, there is a clear need for instructions and agreement on the basic principles of data input. To achieve this, all researchers are trained and given proper guidance. Regular evaluation of the guidelines and feedback ensure that everyone is on the same page. Where possible the guidelines are implemented and enforced in the system itself. Heavy use is made of controlled vocabularies and other validation mechanisms to ensure that any error that can be caught by a machine is indeed caught. But even with training, guidelines,

regular feedback, and machine-driven data validation, inconsistencies in the data will arise. To err is human, after all. Therefore, we have a limited number of editors who are appointed to do a final edit. Only when they have approved a certain record, it will be made available to the general public. In the future, we will keep on using this system. But, where we used to have separate editors per inventory, we are moving towards one team of editors that will need to collaborate to ensure the integrated records are valid for all disciplines. This does require a greater understanding and mutual appreciation for all parties involved. This process in itself has produced valuable insights.

Integrating and streamlining the different databases and editing processes requires more coordination than was previously needed. To capture this need better than was previously possible, a data manager was appointed in 2015.

Apart from coordinating the data entry in the different inventories and managing the quality of the data in them, the data manager has another important role to play. He or she forms the link between the teams that manage the inventories and the information team that engineers the design and development of the information systems involved. To ensure that a clear and integrated vision is maintained at all times, bi-weekly meetings are held. During these meeting status reports are delivered on both development of new features or bugs and data entry or surveying progression. When changes to the data model or the system are wanted, they are brought before the group and the impact on all inventories, disciplines and systems concerned is assessed.

5. SHARING SYSTEMS: FROM BIG BOX TO TOOLBOX

The system we launched in 2009 and have been maintaining and improving ever since was in essence a big box comprising a number of modules. Not only were there core modules to deal with the heritage objects themselves, but also supporting modules that allow us to maintain our thesauri, images and their metadata, biographical data on people related to heritage, etc.

Just as our researchers have transitioned from seeing their respective inventories as being significantly different from their colleagues inventories to grasping the actual similarities between them, our information specialists have come to the realisation that there are synergies that can be achieved with other systems, both within and outside of our own organisation.

Flanders Heritage is currently expanding and reworking its portfolio of digital systems due to major legislative changes. In doing this, it has become clear that the modules currently part of the inventory system could also be of use to other systems being built. Therefore, we are reworking these modules towards standalone systems that can be used by the entire agency, not just by the inventory or the department that maintains the inventory.

While the agency has always been an avid consumer and supporter of open source software, we have only recently started producing open source software in house. Through our involvement in larger heritage communities such as those of the Arches project (Myers et al., 2012) and our contacts with colleagues in neighbouring countries, we came to understand that there's an untapped potential for collaboration when it comes to software. Most of our core systems (such as the inventory) are so specifically tailored towards our own business processes that they are unsuitable to anyone but us. But some of our secondary systems can be of use to other agencies of the Flemish Government, other

authorities in Belgium, and the wider heritage community. Currently, we have released libraries for interacting with the central address database of Flanders (CRAB), libraries for handling SKOS vocabularies (Skosprovider), and an editor for controlled vocabularies and thesauri (Atramhasis), all through our Github page (<https://github.com/OnroerendErfgoed>). In the future, other applications and libraries such as an application for maintaining an image database might be added as well.

6. FROM THE BOOKSHELF TO THE INTERNET: THE CONNECTED PAST

The transition from book to website has opened up opportunities for raising the awareness of the general public about heritage. While traditionally the books would be distributed and marketed through analogue channels, this is no longer wanted or even necessary with the existence of a digital medium. Progressively, this has shifted towards the use of modern day communication strategies. And while an inventory is very much focused on the heritage itself, promoting the inventory and the heritage works best in an interactive fashion, by including the audience and letting them participate.

We have done this with interactive games (we've lost a building who can tell us where it is?), by crowdsourcing work we do not have the time for (taking pictures of items that were added to the inventories when photography was still very expensive), and by having our own researchers or colleagues present their favourite item in the inventory.

Whenever there are marked historic events (e.g. the centennial of World War I or the bicentennial of the Battle at Waterloo), we write newflashes about them and the traces that remain of them in the database. Because all datasets are now sharing a common website, it is very easy to tell stories that transcend the different disciplines.

Throughout the years, we have seen a steady influx of visitors to the website, and have been following the visits to the system since May 2009. While we started with a little over 8000 visits per month, this has steadily risen to about 120.000 visits per month. As we put more effort into dissemination via social media, the number of users that reaches us through these channels rises as well. But by far our largest source of visitors is through search engines. A little over 80% of our visitors reach us through a search engine, predominantly Google (98%). We have always taken steps to ensure that our pages are optimized to be read by search engines and can easily be crawled. Among other things, we have done this by making sure every item has as good a description as possible. Even though we no longer publish books, a large part of our inventory consists of written descriptions of heritage. Our keywords and thesauri are absolutely vital for retrieving information from our inventories, but our written descriptions are vital for conveying our interpretation of the data and sharing the knowledge present in the system.

7. SOME THINGS WILL NEVER CHANGE

As fundamental as all these changes have been throughout the history of heritage data collection and data management at Flanders Heritage, some things never change.

The digital inventory has provided new and exciting opportunities to structure, present, and search heritage data. Questions that were unanswerable 20 years ago can now be answered with a few clicks of a mouse. The systems have changed from paper books

over simple databases to modern day GIS driven websites. But the information the system discloses is the real bread and butter of this work. Paper or screen: the information remains the same.

The constant interaction between the content, the system, and its users keeps on pushing us to rethink our strategies and our day-to-day operations. Heritage professionals and information systems professionals are constantly challenging each other's beliefs and assumptions. Every interaction creates an opportunity to enrich both the system and the way our researchers see our heritage.

This also means that the work is never done. The system is never finished. There is always a data model that needs to be revised or a new feature to be developed. There are always data that need to be added, adapted and revised. There is always a new story to be told about our heritage and communicated to the public. And in this day and age, it is almost tempting to forget that the system is a means, not an end. As heritage professionals, we need to make sure that enough time is left for actually accruing knowledge about the heritage itself. Otherwise, our fantastic big box will turn out to be nothing more than an empty shell.

REFERENCES

Hooft, E., 2011. Bouwkundig erfgoed in kaart gebracht. Methodologie en conclusies van de administratieve actualisatie van de inventaris van het bouwkundig erfgoed in vlaanderen. *Relicta, Archeologie, Monumenten- en Landschapsonderzoek in Vlaanderen* 7, pp. 239-260.

Hooft, E. and Verwinnen, K., 2008. Onderzoeksbalans Onroerend Erfgoed Vlaanderen. B.2 Inventarisatie van bouwkundig erfgoed. Vlaams Instituut voor het Onroerend Erfgoed, Brussel. <https://onderzoeksbalans.onroerenderfgoed.be/onderzoeksbalans/bouwkundig/inventarisatie> (24 Mar. 2015).

Myers, D., Dalgity, A., Avramides, I. and Wuthrich, D., 2012. Arches: An open source gis for the inventory and management of immovable cultural heritage. In: Ioannides, M. et al (eds), *Progress in Cultural Heritage Preservation: 4th International Conference. EuroMed Conference 2012, Lemessos, Cyprus, October 29- November 3, 2012, Proceedings*, pp. 817-824.

Van Daele, K., Meylemans, E. and De Meyer, M., 2004. De Centrale Archeologische Inventaris: een databank van archeologische vindplaatsen. In: *CAI-I, De opbouw van een archeologisch beleidsinstrument*. IAP-Rapporten 14, Instituut voor het Archeologisch Patrimonium, Brussel, pp. 29-48.

Van Lindt, P., Kennes H. and Hooft, E., 2006. Voorbeeldproject georeferentie inventaris bouwkundig erfgoed. In: *Vlaams Instituut voor het Onroerend Erfgoed. Een nieuw perspectief voor erfgoedonderzoek*. Vlaams Insituut voor het Onroerend Erfgoed, Brussel, pp. 58–62.