Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

J.Y. Blaise, I. Dudek, G. Saygi
 UM CNRS/MC 3495 MAP
Our research area:  **Data science**  Heritage studies

Taking inspiration from InfoVis methods and practices in order to renew (or at least complement) the way historical evidence can be analysed, cross-examined, and given to reason on.

- A strong concern for the **time** parameter (chronology but also cycles, rhythms, etc.)
- A strong concern for the **quality** of the data (lacks, whole range of uncertainties)
Our research area:  Data science  Heritage studies

Taking inspiration from InfoVis methods and practices in order to renew (or at least complement) the way historical evidence can be analysed, cross-examined, and given to reason on.

- A strong concern for the time parameter
- A strong concern for the quality of the data

A focus on the architectural heritage (but not only).
Our job: pulling together evidence and hints in order to read or re-read data and information sets

NOT in an attempt at pushing to the fore and propagating false beliefs (virtual reconstructions)
Our job: pulling together evidence and hints in order to read or re-read data and information sets rather in an attempt at laying on the table « what we really know, », i.e. at formalising and cross-examining pieces of data and information >> enhancing relations

Successive documented “versions” of an edifice as time passes by.
Either: a or nothing

If $\text{evt}$ then $a$ or $b$

Unproven filiation link

Typically, lootings following an assault on the city.

Typically, a fire, with unclear consequences

J.Y Blaise, I.Dudek, G.Saygi, UMR CNRS/MC 3495 MAP

Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

J.Y Blaise, I.Dudek, G.Saygi, UMR CNRS/MC 3495 MAP

Context (laws, epidemics, floods, etc.)

Military events (occupations, sieges, etc.)

Correlation: observing the impact of the Czech army’s presence in Krakow on the building activity

Our research area: Data science Heritage studies

Taking inspiration from InfoVis methods and practices in order to renew (or at least complement) the way historical evidence can be analysed, cross-examined, and given to reason on.

- This presentation may well not be fully inline with the scope of DSAA, but research paths such as ours imply taking the risk of being the ugly duckling.
Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

J.Y Blaise, I.Dudek, G.Saygi, UMR CNRS/MC 3495 MAP

[outline]

Context and motivation

Range of issues raised

A focus on the visualisation effort

Yet to come (i.e. shortages)
A research programme that introduces a number of constraints:

> Numbers - yet « small data »:
- over 1200 edifices (rural chapels, southeast of France)
- Over 3300 web pages documenting these edifices

1265 edifices
16 unlocalised edifices
122 ill-localised edifices (textual hint)
1092 chronological hints
263 contradictory chronological hints
468 edifices undated
3438 web resources
624 information providers
52 edifices undocumented (no web resource available)
A research programme that introduces a number of constraints:

> numbers:

> Nature of the objects: *Minor heritage*

- NOT likely to be documented by public authorities or academics in a normative, consistent way
A research programme that introduces a number of constraints:

> numbers

> Nature of the objects: *Minor heritage*

- NOT likely to be documented by public authorities or academics in a normative, consistent way

- We rely on (often local-based) enthusiasts
A research programme that introduces a number of constraints:

> **numbers**

> **Nature of the objects**: *Minor heritage*

> **Profile of the information providers**

- Variety in motivation, and in practices: the nature of the data they are likely to publish is far from being easy to foresee.

- A whole range of uncertainties
A research programme that introduces a number of constraints:

> numbers

> Nature of the objects: **Minor heritage**

> Profile of the information providers

> Density of hints per edifice

- Variety in the amount of data available for this of that edifice, including inside a small perimeter.
A research programme that introduces a number of constraints:

> numbers

> Nature of the objects: **Minor heritage**

> Profile of the information providers

> Density of hints per edifice

> Intensity of activity of information providers

- a typical *citizen science* pattern (10/90 ratio - 10 % of the contributors do 90 % of the job)

- also, a variety of “behaviours” from those engaged in a systematic “survey” of such edifices to people only interested in the history of their village or valley.
A research programme that introduces a number of constraints:

> numbers

> Nature of the objects: *Minor heritage*

> Profile of the information providers

> Density of hints per edifice

> Intensity of activity of information providers

> Finally, considering the numbers above, at this stage of the research we do not analyse the information, but simply pull it together, and echo initiatives of citizens (a cornerstone of a sound and balanced relation between academics and information providers)
A research programme that introduces a number of constraints – so what is the motivation?

> Get a better understanding of what can be gained, at scientific (and cultural) level from this information mash-up

> Our aim was not to document and analyse thoroughly the corpus, but to somehow weigh, in terms of methodology, the potential impact of citizen-birthed data and information sets on the documentation and analysis of minor heritage items.

> A web portal used as a tool to foster discussions, share data, experiment contribution modalities, and promote visual reasoning.

http://www.map.cnrs.fr/territographie/
The paper reflects the fact that we have been pulling several wires all along the research.

A brief overview of the research issues raised, and a focus on the visualisation effort.

> A grid of uncertainty factors: do such data sets contain factors of uncertainty that are met in “classic” historical research?

<table>
<thead>
<tr>
<th>determinacy</th>
<th>Whether the value of a variable is known at all or not.</th>
<th>[18]</th>
</tr>
</thead>
<tbody>
<tr>
<td>An edifice that cannot be localized (only mentioned in archives, without hints on its position), or that cannot be dated.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>credibility</th>
<th>Judgment made by the human consumer of the information about the information source.</th>
<th>[19]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association of a craft and a territory basing on the sole recollections of a witness.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>approximation</th>
<th>Attempt to come close to measuring or describing a phenomenon [...].</th>
<th>[19]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measuring the orientation of an edifice bases on the presumption that the nave is actually straight, and the apse unambiguously positioned.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>incompleteness</th>
<th>The idea that the observed evidence is likely to only be a small portion of the whole.</th>
<th>[21]</th>
</tr>
</thead>
<tbody>
<tr>
<td>None of the collections we handle is complete, hence the necessity to be cautious in any interpretation. Furthermore, as mentioned by [19], the unidentified unknowns are the worst kind of missing information, and one of the sub-goals of a citizen science approach to minor heritage can be to try and diminish the amount of unidentified unknown.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>interrelatedness</th>
<th>Source independence from other information.</th>
<th>[21]</th>
</tr>
</thead>
<tbody>
<tr>
<td>When two e-sources make a common statement, yet without quoting each other or the common initial source they based on.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>currency</th>
<th>Temporal gaps between occurrence, info collection &amp; use.</th>
<th>[21]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal gaps between the period of use of a farming tool, and the moment when it was collected.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>multivocality</th>
<th>When several hints appear as contradictory.</th>
<th>[18]</th>
</tr>
</thead>
<tbody>
<tr>
<td>The literature sometimes mentions inconsistency or disagreement [21] to name such imperfections – a typical occurrence is opposite contradictory dates given for an event.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The paper reflects the fact that we have been pulling several wires all along the research.

A brief overview of the research issues raised, and a focus on the visualisation effort.

> A set of contribution modalities needing experimentation – What would be the best-suited technological environment for information providers?
The paper reflects the fact that we have been pulling several wires all along the research. A brief overview of the research issues raised, and a focus on the visualisation effort.

> An attempt at profiling information providers – who are they? Is there a way to group them by clusters? According to which division lines?

> If there are profiles, can they be aligned with “types” of data sets that a given profile is likely to provide?

*e.g.* are those who provide dates of construction of an edifice (linear time) the same as those who tell us when a religious parade is organised for that edifice (cyclic time).
The paper reflects the fact that we have been pulling several wires all along the research.

*A brief overview of the research issues raised, and a focus on the visualisation effort.*

> An attempt at profiling information providers – who are they? Is there a way to group them by clusters? According to which division lines?

> If there are profiles, can they be aligned with “types” of data sets that a given profile is likely to provide?

*e.g.* are those who provide dates of construction of an edifice (linear time) the same as those who tell us when a religious parade is organised for that edifice (cyclic time).

Relations between profiles of information providers and types of dissemination activities they are likely to participate in (or to organise).
The paper reflects the fact that we have been pulling several wires all along the research.

*A brief overview of the research issues raised, and a focus on the visualisation effort.*

> How do information providers react to an academic or “official” research targeting a corpus they were the first to investigate?

> What would they expect in a move from “crowdsourcing” practices (citizens called in to back up the initiative of academics) to citizen science practices (where they act as initiators)?

Maps representing the spatial distribution of edifices documented by a given information provider and quoted in the system, produced on the fly as a service for information providers.

Top, the information provider proposes a systematic “survey” of rural chapels
Bottom, the information provider documents various aspects of a given alpine valley, beyond the corpus of rural chapels
The paper reflects the fact that we have been pulling several wires all along the research.

**A brief overview of the research issues raised, and a focus on the visualisation effort.**

> How do information providers react to an academic or “official” research targeting a corpus they were the first to investigate?

> What would they expect in a move from “crowdsourcing” practices (citizens called in to back up the initiative of academics) to citizen science practices (where they act as initiators)?

Visual support of a series of workgroup discussions organised in the flow of the project so as to get feedback from information providers.
The paper reflects the fact that we have been pulling several wires all along the research.

A brief overview of the research issues raised, and a focus on the visualisation effort.

> Finally – what kind of lesson can we learn about such heritage items basing on such data and information sets? Will we manage to identify an added-value?

An illustration through two examples, concerning the edifices collection:

- One that focuses on what we can learn about the data quality
- One that focuses on what we can learn about the edifices themselves
To start with, some definitions and a clarification concerning the data considered.

> Spatial data (edifices only)

Geographical positions

*Longitude, latitude*

Hierarchical structure

*edifice, commune, département (~200 to 300 communes)*

Altitude

*in meters, above sea level*

Orientation

*in theory, Christian churches are “oriented” (the apse to the east, the Altar there)*
To start with, some definitions and a clarification concerning the data considered.

Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up.

J.Y Blaise, I. Dudek, G. Saygi, UMR CNRS/MC 3495 MAP

Funded by the Order of the Knights Templar, at a time still being discussed.
Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

J.Y Blaise, I.Dudek, G.Saygi, UMR CNRS/MC 3495 MAP

Oct.2018 Torino (IT)
To start with, some definitions and a clarification concerning the data considered.

> Temporal data

Time anchors - a concept developed as the actual hints we meet may refer to the construction of the edifice itself (“built in the 17th century”), but also to its “first mention in archives”, etc. Basically the association of the edifice to a verbal or quantitative expression of a date.

e.g. “probably built in 1766” ; “beginning of the 17th century”, etc.

\[ \text{Expression} \rightarrow \text{rule} \rightarrow \text{YYYY Integer(s)} \]

\[ 1 \text{ edifice} \leftrightarrow 0 \text{ to } n \text{ Time anchors} \]

Temporal cycles of use, changes, etc.
To start with, some definitions and a clarification concerning the data considered.

> Thematic layers

Toponymy (alternative names)

Architectural analysis (e.g. shapes, components)
To start with, some definitions and a clarification concerning the data considered.

> Thematic layers

**Toponymy (alternative names)**

**Architectural analysis**
(e.g. shapes, components)

**Work practices in traditional crafts** (e.g. who, learning, places of practice, etc.)
Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

Data quality: does it depend on the commune considered? (partial view)

Colours: départements

Rectangles: communes

Jausiers [6]
Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

J.Y. Blaise, I. Dudek, G. Saygi, UMR CNRS/MC 3495 MAP

Oct. 2018 Torino (IT)

[the visualisation effort]

Vertical lines: edifices

Chapelle des Payans (227)

The localisation information (where?)
The orientation information (how?)
The chronology information (when?)

The information exists
Contradictory information found
No data

[Image of visualisation with dots representing information and labels indicating where, how, and when information is available or contradictory or missing.]
Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

J.Y Blaise, I.Dudek, G.Saygi, UMR CNRS/MC 3495 MAP

The visualisation effort

An overall picture, that can then be filtered

- The information exists
- Contradictory information found
- No data
A focus on dates alone
Pattern analysis: « all correct »
Pattern analysis: « all wrong »
about the edifices - Collection reading: Is the orientation rule (apse facing the east) really applied to small chapels? Until when? Does the relief or the altitude impact its application?

One rectangle = one edifice

Position around the compass: Orientation

Colours = département
Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

J.Y. Blaise, I. Dudek, G. Saygi, UMR CNRS/MC 3495 MAP

[the visualisation effort]
Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

No data
The same, but this time showing dates: the darker the older. Outlined in red: non-dated edifices.
A tendency of older edifices to be better in line with the rule but also an inclination of later builders for turning the apse towards the north.
Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up

De-correlates the orientation and altitude parameters (in other words, denies the common sense belief that because of stronger relief constraints builders are more keen to step out of the rule).

The same, but this time showing altitudes: the darker the higher.
An ongoing research.

> Understanding and foreseeing the consequences of relying on citizen-birthed data when trying to document and analyse minor heritage items.

> First outcomes: data sets, visualisations, feedback on the expectations and practices of information providers.

> More needs to be done, in particular in terms of analysis of the information providers’ profiles.
Conclusions

> Understanding and foreseeing the consequences of relying on citizen-birthed data when trying to document and analyse minor heritage items.

> First outcomes: data sets, visualisations, feedback on the expectations and practices of information providers.

> More needs to be done, in particular in terms of analysis of the information providers’ profiles.

Relations of IP profiles to publication practices.

The circle on the right represents “publication of raw data” – only three profiles do consider the publication of such data sets, and two of them occasionally only (dashed line).
Two key observations:

- Minor heritage, because of the data quality / heterogeneity issues that are met, is a challenging area of research for both information sciences and heritage sciences.

- An unforeseen picture of the situation, born in the flow of the project: the main challenge today may not be this of developing crowdsourcing practices, harvesting « new data » - but may well be to oppose the volatility of the pieces of information and knowledge « already there » - somehow a paradox given the technology at hand.

http://www.map.cnrs.fr/territographie