



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

Jean-Yves Blaise, Iwona Dudek, Gamze Saygi

► To cite this version:

Jean-Yves Blaise, Iwona Dudek, Gamze Saygi. Citizen contributions and minor heritage: feedback on modeling and visualising an information mash-up. The IEEE International Conference on Data Science and Advanced Analytics (DSAA) 2018, Oct 2018, Torino, Italy. halshs-01887679

HAL Id: halshs-01887679

<https://shs.hal.science/halshs-01887679v1>

Submitted on 4 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Marseille, France

e-mail: [jean-yves.blaise\(iwona.dudek\)\(gamze.saygi\)@map.cnrs.fr](mailto:jean-yves.blaise(iwona.dudek)(gamze.saygi)@map.cnrs.fr)

Abstract—The citizen science paradigm and the practices related to it have for the last decade called a wide attention, beyond academics, in many application fields with as a result a significant impact on discipline-specific research processes and on information sciences as such. Indeed, in the specific context of minor heritage (tangible and intangible cultural heritage assets that are left aside from large official heritage programs), citizen-borned contributions appear as a major opportunity in the harvesting and enrichment of data sets (notwithstanding data quality and heterogeneity issues). In parallel, it seems we have today reached a moment when the acquisition and analysis of spatio-historical information appears “easier” since citizens are seen as potential (and legitimate) sensors. But is it really “easier”? And if so, at what cost? Having a closer look on practical challenges behind the curtain can avoid turning the above mentioned opportunity into a lost one. In this contribution we present a research initiative that aims at better circumscribing the difficulties one has to foresee when wanting to harvest and visualize pieces of data on minor heritage collections, and then to derive from them spatial, temporal, and thematic knowledge. The contribution focuses on three aspects: a short analysis of citizen contributions in the context of minor heritage, a description of the case study and of the data modeling bottlenecks we are facing, and an exemplification of the visual analysis solutions we experiment in order to portray and question our understanding of collections. The case study acts as a test bench helping to investigate data harvesting and modeling challenges, as well as potential added-value of the visualization step.

Keywords-Spatio-historical data modelling; Citizen Science; Information visualization; Knowledge Discovery; Research Methodologies; Minor Heritage

I. INTRODUCTION

The notion of minor heritage refers to material and immaterial forms of unprotected cultural heritage, often products of rural societies. They are genuine signs of craftsmanship, culture and traditions, and important components of the life of former generations. Nevertheless, quite often minor heritage slips through the large heritage programs or documentation initiatives, and therefore there is a lack of appropriate knowledge, and an unconsciousness of its value.

In that context citizen-borned contributions appear as a major opportunity in the harvesting and enrichment of data about heritage assets (tangible or intangible, real-estate or

movable). However actors potentially concerned – from active citizens to scientists or collection holders - should not overlook the difficulties created by the nature of the hints one will face in that application field, for instance in the verbalization (e.g. “probably in 1666”) or in the precision of the data (e.g. “alongside a former riverbed”).

More generally, minor heritage is concerned with memories, with history, and history deals by essence with ill-defined spatial, temporal and thematic data. Generally speaking, renewing the perceived value of such assets can help cultural actors (from citizens to “officials”) in their effort to preserve the assets and assess their significance. Just as in T.Pratchett’s Discworld gods get stronger when people believe in them, or weaker when they cease to do so, the understanding and preservation of minor heritage assets relies on the will of people to get engaged.

It is therefore important, along with foreseeing the potential benefits for scientists, to examine how trendy “crowdsourcing” or “citizen science” initiatives can contribute in a sustainable manner to the acquisition and analysis of heritage information sets, as well as to support actors already engaged in assessing the cultural value of minor heritage assets.

And indeed over time the “citizen science” concept tends to get blurred, and who exactly is meant when using the terms people, or citizens, can be a source of misunderstanding. Is a contributor to a crowdsourcing platform primarily a citizen, an enthusiast engaged in this or that field, or primarily “just a web user”?

As will be shown, the heart of the data we handle originates from citizens, who get involved – with or without relation with academics – and actually produce and publish on the net large data sets. They, indeed, are citizens as well as web users; but they are primarily as far as our research is concerned local actors strongly engaged in Minor Heritage documentation, may it be for professional reasons or not. Accordingly, in the context of this research, we shall call them simply “Information Providers” (IP). Briefly said, the involvement of Information Providers appears as promising, yet demanding, in particular due to their heterogeneity:

- variety in terms of motivation, ranging from personal, private involvement of Heritage enthusiasts to a broad public sector set of actors (local communities engaged in cultural tourism for instance),

- variety in terms of analytical biases, ranging from systematic, inventory-like approaches (privileging here a thematic entry - “oratories”, “old tools” - or there a geographic entry - “all about my village”) to subjective selections based on the perception, emotions of the information provider,
- variety in terms of work processes, in relation with the familiarity of the information providers with information technologies at large, and notably with web publishing platforms.

Accordingly, our research aims at better circumscribing the difficulties one has to anticipate when wanting to harvest and visualize those pieces of information, and to derive from them pieces of Knowledge. To do so, we have pulled together three collections, that each push to the fore practical data acquisition, modelling and interpretation challenges. Generally speaking the research pushes to the fore a number of open issues:

- Citizen science in the context of minor heritage: *in what does that paradigm renew the way one can expect to enrich or analyse minor heritage assets?*
- Minor heritage in the general context of heritage sciences: *what kind of general knowledge can be derived from such items and collections, from partial, often unverifiable hints?*
- Information visualization (InfoVis) [1] and Visual Analytics [2] methods as means to foster reasoning and sensemaking in the context of large, but citizen-borned, ill-defined information sets: *to which extent do these method apply?*
- Open-source geovisualisation solutions in the context of strongly heterogeneous data (movable vs. real-estate, tangible vs. intangible): *what solutions can one rely on today, and how can the temporal dimension be efficiently visualized?*

But this paper is essentially about feedback: we focus on *practical challenges* we have faced in terms of data modeling and visualization, and introduce three main contributions.

Section 2 positions the research context with an emphasis on how citizen science practices meet minor heritage collections, and presents the strategy we have experimented in order to build mutually beneficial relations between Information Providers and the research program. **Section 3** details the case study, pitfalls and choices made to cope with *imperfections* in the information [3]. In that section we identify and exemplify *factors of imperfection* that we met in this very specific context of heterogeneous minor heritage documentation using e-sources. In **Section 4** we present some of the visual solutions we have experimented in order to support analytical tasks in such information sets, and highlight the gain of insight they foster. Section 5 gives an overview of the evaluation strategies we have implemented up to now in order to measure the potential impact of the visualisations produced, as well as to get feedback from information providers and interpret it. Finally, we shortly sum up our main findings at this stage, and list some of the limitations and bottlenecks still ahead.

II. RESEARCH CONTEXT

At first glance this research could be seen as another contribution to this general and powerful move towards Digital Atlases – in the legacy of archaeological or architectural atlases, i.e. systematic, geo-referenced collections of pre-formatted information sets. Such initiatives encompass broad concerns that range from graphic semiology [4] to compatibility with standards (or more generally interoperability issues) [5] [6]. In that general context of “formatting and interfacing heritage data sets” this research however builds on very specific choices, strategies, and issues that need to be made clear:

- The data we handle is not pre-formatted, and therefore falls outside of the classic “official records” approach to heritage items. We use heritage items as nodes (located in time and space) interlinking strongly heterogeneous e-sources.
- The focus is put in this research on the potential added value of cross-examining though visual means such heterogeneous, and “imperfect” data sets.
- We are concerned with the way the information provided by citizens, as it stands today, can feed the analysis of heritage items. This self-imposed constraint delineates the type of findings expected: getting a better understanding of the information sets themselves, of what analytical processes they are likely to support, and not necessarily getting a better understanding of the heritage items themselves.

As a consequence this research claims no contribution to the creation or interfacing of heritage atlases – its scientific context is the impact of citizen-borned information sets on scientific processes. And indeed with the emergence of information communication technologies citizens are becoming an increasingly important information source for diverse domains. Pioneering studies are done in geographical/spatial information thanks to open infrastructures like OpenStreetMap [7] and free web-mapping services facilitate an implementation of the “citizens as sensors” mantra [8], with contributors generating content, and enhancing geographical information by pointing out their local particularities. In a decade time, crowdsourced geographical information took various shades of application - e.g. voluntary geographical information, contributed geospatial information, user - created content [9], all in all, replacing the top-down tradition of geographical information production [10]. But this move renews challenges such as inconsistency issues, data validation and quality assessment requirements [11, 12].

In parallel typical citizen contributions for cultural heritage focus on basic tasks like annotating and tagging photos of archival material with keywords [13], transcription of historical texts [14], attributing longitude and latitude of objects by dynamically creating markers on the map [15]. But what one needs to harvest in the context of our application field is not limited to the above indications: what we need to harvest are personal memories, records of individual experiences, self-knowledge about a time gone by. In other words, the information we need to uncover is more

detailed yet less precise; it is often subjective, unverifiable, hence a questioning of the above contribution modalities. And even if we base on the assumption that contributors are good-willing and trustworthy, an experience reported in [16] shows we ought to be cautious. In that experience a group of people was asked to say how they were informed of the 1986 Challenger space shuttle accident, a short while after the accident itself, and then again 20 years later. Differences in the reports talk for themselves: recollections are significantly altered by time.

Yet uncovering personal, individual self-knowledge in citizen contributions is crucial if we want to preserve and share an understanding of minor heritage. Furthermore it does correspond to a trend that originates from citizens themselves who are today uploading self-knowledge and past experience via personal blogs.

In other words the issue of how to record, share and interpret such contributions is raised. As an answer, we use the case study as an opportunity to test out alternative contribution modalities, that target individuals' self-knowledge rather than solely focusing on collecting or decrypting massive information sets, and that can contribute to a better assessment of what imperfections are likely to be met.

More than ten years ago, M.F.Goodchild [8] coined the terms "citizens as sensors", paving the way for the Volunteered Geographic Information (VGI) research community, while the Galaxy Zoo experiment and its extension the Zooniverse platform introduced a more generic concept referred to as people-powered research (and the latter has indeed been of use in Heritage Studies since then). Yet over time shades of that generic concept have clearly emerged [17]: crowdsourcing practices where citizens get engaged in order to support research activities initiated and conducted by academics, citizen science projects launched by citizens themselves, and finally collaborative science where citizens and academics co-define the research topic. With regards to those shades, our contribution primarily falls within the citizen science paradigm: the engagement of Information Providers, and therefore the data we base on, obviously existed prior to our research.

Yet we have initiated a series of workshops with IPs in order to move towards a collaborative approach by co-narrowing research questions. But in order to do so one has to be pay attention to potential misunderstandings between academics and non-academics, and in particular to clearly assess mutual benefits. Our strategy has been to try and demonstrate potential added-values on both sides.

This was done as a first step by showcasing the accomplishments of IPs in three different ways: each leaf of the platform (i.e. pages presenting one particular heritage item, reached after user selections) contains links to the information providers' platforms; one of the search modes is in fact a list of IPs, with for each of them a list of items they contribute to document; and a specific visualisation highlights the intensity of an information providers' activity on the territory under scrutiny (Fig. 1).

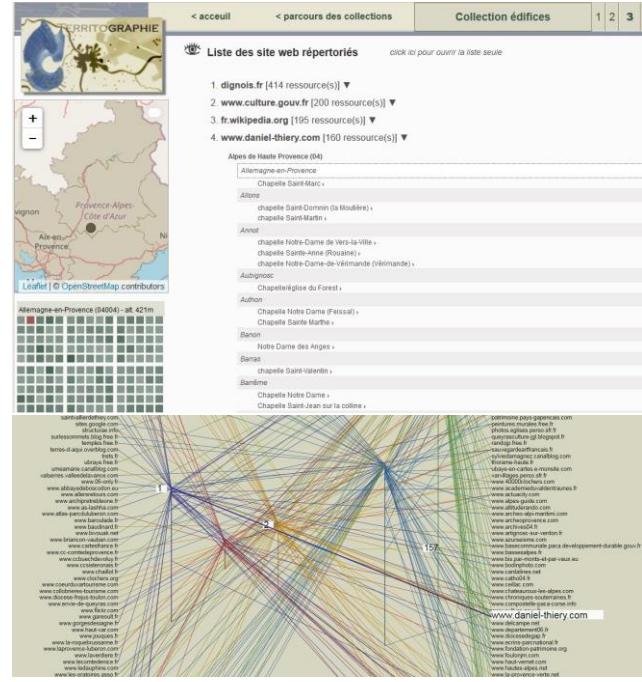


Figure 1. Top, a browsing mode that showcases the information providers works – list of all edifices documented by an information provider (here daniel-thierry.com, in-depth historical analyses authored by a local expert on a personal basis). Bottom, a visualisation in which the geographical distribution of edifices that a given information provider contributes to document is shown. Vertical lines correspond to the six départements in the region (intermediate administrative level). Oblique lines connect a given information provider to a “number of e-sources provided” for each département. In this example daniel-thierry.com appears as quoted 157 times in one département (green), but only once or twice in two others : a clear marker of this information provider’s area of concern, and type of practice.

We then developed customized services for each information provider: on-the-fly production of three visualisations that show the information sets corresponding to one information provider (Fig. 2, Fig. 9, Fig 11).

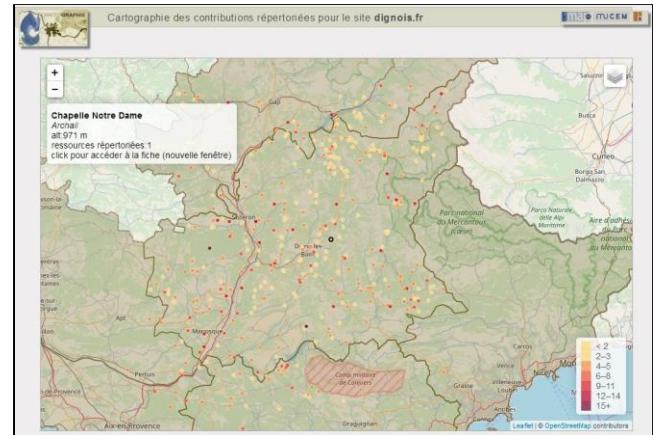


Figure 2. Cartography customised to show the pieces of information produced by one specific Information Provider (here dignois.fr). Other customised services – chronography and density / altitude analysis visualisations - are discussed in section 4.

One should not however overlook the difficulty of aligning this strategy with the actors. Typically what we view as “services” can be totally irrelevant for this or that type of IP. Hence a necessity to try and better understand profiles of IPs, their expectations, and their work processes. At this stage of the research we identified eleven profiles (e.g. community builders, bloggers, collectors, local actors, etc.) and aligned these profiles with behaviours or work practices (type of content published, thematic scope, geographical area, etc.). This assessment of ways of doing proved helpful in the preparation of workgroup discussions with IPs (see section V) in order to unveil tendencies in the material they publish, or in their editorial choices (Fig. 3).

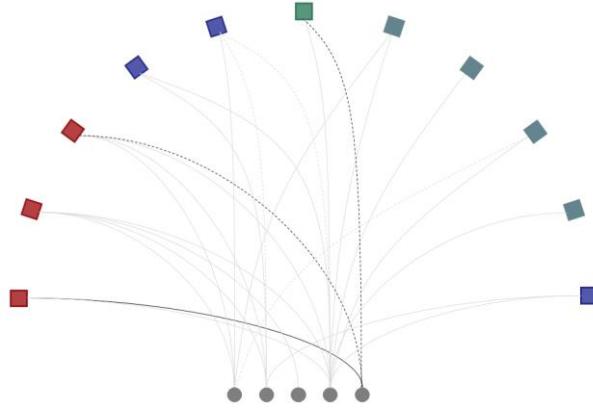


Figure 3. Relations of IP profiles (represented as coloured squares - colours correspond to “types” of actors - e.g. red for parapublic sector, blue for NGOs, etc.) to publication practices. The circle on the right represents “publication of raw data” – only three profiles do consider the publication of such data sets, and two of them occasionally only (dashed line).

III. THE CASE STUDY

Our case study concerns the territory of the PACA (Provence Alpes Côte d’Azur) region, one of France’s 13 metropolitan regions, composed of 958 communes (smallest administrative layer). These communes are distributed in 6 départements, an intermediate administrative level. Communes vary in size (area and population, in proportion 1:1300 and 1:3449, 2013 data). We pulled together three collections consisting of : (i) 1217 rural chapels, (ii) 360 traditional farming tools from the MuCEM Ethnological museum (Musée des Civilisations de l’Europe et de la Méditerranée), and over 200 traditional crafts and professions (Fig. 4). The idea is to try and cross-examine components of the collections and to assess visually spatial, temporal, and semantic relations.



Figure 4. An illustration of the items in the various collections.

A. Harvesting the data

Although the collections are different, we extract a set of common features corresponding to the spatial dimension (classic point location, but also relations between locations as in commercial exchanges), to the temporal dimension (both linear chronology and potential cyclic behaviors), and to thematic layers underlining potential spatial and temporal concurrences across the collections.

Table 1 sums up the data harvested (or available) for each collection.

TABLE I. THE DATA - SPATIAL, TEMPORAL AND THEMATIC DIMENSIONS.

	Spatial	Temporal	Thematic
edifices	Location (long, lat), altitude and orientation ¹ . The notion of “spatial data” encompasses the 3D shape of the edifice – one of our fields of concern, but a topic we do not mention in the context of this paper.		
	Time anchors: a concept developed as the hints we meet may refer to the construction of the edifice itself (“built in the 17 th century”), but also to its “first mention in archives”, etc.. A time anchor is basically the association of the edifice to a verbal or quantitative expression of a date. Other temporal indications recorded are temporal cycles of use (seasonal use for instance) and changes that occurred along the history of the edifice.		
	Thematic layers range from architectural analysis (e.g. shapes, components) to records of experiences such as pilgrimages, votive festivals, or simply travel diaries.		
farming tools	An item is, when possible, associated with a geographical container – commune or département.		
	Two indications are available: the recording date (year of acquisition by the Museum) and when possible the period of use of the item (No indication on the cycles of use).		
crafts and professions	Items are grouped by categories that act as thematic layers. Scarce indications are available on the “instructions for use”, on the making of the items, on the way they were stored and preserved – these are the pieces of information the research primarily expects to harvest through citizen contributions.		
	Association with one or several containers (communes) and when possible with a specific location (market, etc.)		
	Recording of periods of practice (question: “when was this craft present”)		
	Thematic layers include hints on the practices themselves, on their naming (etymology, alternative names), on tools and materials concerned, and when relevant a qualitative description of the spatial layout (mobile stand, workshop, etc.)		

The data was harvested primarily from online e-content concerning the edifices and the traditional crafts collections. For the edifices collection 3054 web pages are recorded, along with an attribute stating the “nature” of the information provider: public service, citizen contribution (from non-specialized blogs to heritage supporters’ web sites), community building initiatives, etc. Fig. 5 gives a broad idea of the proportion of “citizen-birthed” information.

¹ Orientation of the edifice’s nave, i.e. of a vector running from the porch to the apse, by convention counted from the North.

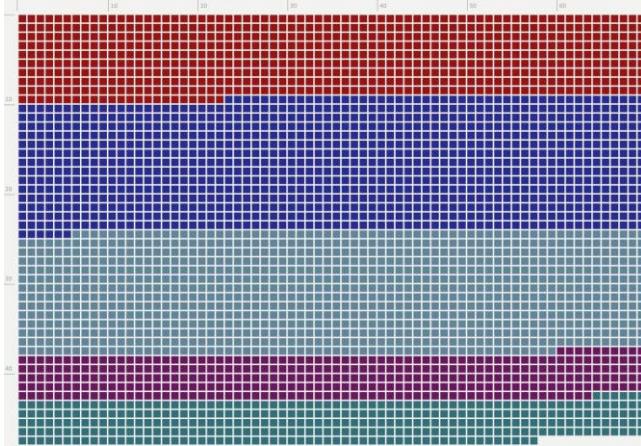


Figure 5. Information providers' profiles: each square correspond to an e-source. Green, blue and light-blue squares correspond to citizen-borned information, red to public services, purple to commercial site.

Concerning the farming tools collection most of the data was already available in the MuCEM's records, however e-sources were harvested in order to complement definitions or to associate items with recurrent cyclic events (fairs typically).

For the traditional crafts and professions collection a set of e-sources was also collected that document the crafts in general terms or as they developed on the territory we are interested in. But here a traditional bibliography complements the information available, and multimedia content (still images, videos and sounds) are referenced that give a more "ethnographic" colour to the data harvested. On the overall 1139 e-sources are referenced. In both the above cases the data was harvested manually – a limitation in a sense, but a necessity in the early stages of the research in order to grab a fine-grain understanding of the nature of the data.

B. Modelling the data: pitfalls and choices

Defining and verifying the application of a robust *observation protocol* is one of the most prominent difficulties when harvesting crowdsourced data, whatever object of study is concerned. In this research we base on existing data sets, with observation protocols that are hardly described, when not simply non-existent. Our focus is therefore put on the data *as it stands*, and our attempt has been to try and list the consequences, in terms of reliability, of using such data sets.

Classifying factors of imperfection in spatio-temporal or historical data has been a recurrent research topic over the past years, sometimes within the boundaries of a scientific discipline [18][19], sometimes in more interdisciplinary settings [20]. In this section we borrow notions and terms from the scientific literature and from previous research, however some of these terms and notions remain today discussed. We therefore propose in the table below a series of definitions and references in order to avoid ambiguities, along with a series of "real-life" examples of where and how they are present in our case study. We make no claim that this list is comprehensive, yet it does pinpoint some key

factors we came across, in a practical manner, and we consider it a significant part of this "feedback" our contribution is about.

TABLE II. FACTORS OF IMPERFECTION: LEXICON, REFERENCES, AND EXAMPLES OF OCCURRENCES IN THE DATA SETS.

determinacy	Whether the value of a variable is known at all or not.	[18]
<i>An edifice that cannot be localized (only mentioned in archives, without hints on its position), or that cannot be dated.</i>		
credibility	Judgment made by the human consumer of the information about the information source.	[19]
<i>Association of a craft and a territory basing on the sole recollections of a witness.</i>		
approximation	Attempt to come close to measuring or describing a phenomenon [...].	[19]
<i>Measuring the orientation of an edifice bases on the presumption that the nave is actually straight, and the apse unambiguously positioned.</i>		
incompleteness	The idea that the observed evidence is likely to only be a small portion of the whole.	[21]
<i>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.</i>		
interrelatedness	Source independence from other information.	[21]
<i>When two e-sources make a common statement, yet without quoting each other or the common initial source they based on.</i>		
currency	Temporal gaps between occurrence, info collection & use.	[21]
<i>Temporal gaps between the period of use of a farming tool, and the moment when it was collected.</i>		
multivocality	When several hints appear as contradictory.	[18]
<i>The literature sometimes mentions inconsistency or disagreement [21] to name such imperfections – a typical occurrence is opposite contradictory dates given for an event.</i>		
accuracy	Difference between heuristic & algorithm.	[22]
<i>Transferring a qualitative indication such as "during the spring" to a given numerical interval.</i>		
imprecision	Inexactness of measurement.	[21]
<i>The value recorded for an edifice's altitude depends on with what instrument / under what climatic conditions the survey was carried out: a non-systematic protocol may alter the precision of the data.</i>		
lineage	Conduit through which info passed (number of steps).	[21]
<i>A typical example is the recording of farming tools: the information available in the records today results from a process that involved at least the donor of the object and the museum expert, and potentially ancestors of the donor, and contemporary successive curators.</i>		
periodization	Dating of a fact by reference to another one.	[20]
<i>Periodization should be interpreted in a broad sense, from examples such as "in the early middle ages" or "rebuilt after the revolution" to natural facts and phenomena such as "during the crops".</i>		

subjectivity	Amount of private knowledge or heuristics utilized.	[22]
<i>There is no reason to think that the way a craft was organized in one location by one craftsman as it is reported by him or his descendants is an objective testimony of what that craft is, yet we rely on such subjective hints to picture it.</i>		
likelihood	At best, a stopgap verbalization in inference making.	[20]
<i>Wordings such as “could have been built shortly after the Wars of Religion”; or “livestock markets probably took place in the open land close to the river bed”.</i>		

Most of these factors are present in the three collections, and a workable approach to “reduce” imperfections is not always within reach. It has to be said clearly that anyway in Historical Sciences reducing imperfections is in fact not a good idea at all: disagreement and open questions are full-blown components of the reasoning process. Our approach bases on the vision that what we need to convey about the data is not assertive interpretations but hints the way they are, yet made more readable, and sharable. We sum up below on one hand choices made in the modelling of the data, and what we view as open issues, lacks and shortcomings, concerning both the spatial and temporal dimensions.

C. The spatial dimension

The baseline geographical data recorded in the edifices collection are a position (longitude, latitude), an altitude, and an orientation. At this stage we defined a list of values (numerical scale) used to “tag” data with regards to determinacy and credibility factors (is it known at all? From what source?). The scale is used in the recording of the data, and at visualization time in the graphic semiology. Yet this does not solve the problem extensively. Edifices for which the location is not known at all are tagged as belonging to a container (a commune) - but what if even that indication is unavailable?

For the crafts collection we record two indications: association of a craft with one or several containers (communes) and qualitative description of the spatial layout. But in the case of occupational travelers, a category that encompasses herdsmen moving to mountain pasture areas and travelling salesmen, we still need to propose a data model that would allow the recording of itineraries. This is not a trivial issue: recording only the start and end location is bland from the scientific point of view, adding “stop points” along itineraries (a watering place, a fair, a rest place) could make sense but leaves vast unknowns since an itinerary is hardly composed of straight lines. What is more, temporal aspects of the displacements are also very significant: a series of (Space, Time) tuples could therefore better match the reality of what we need to learn.

Finally for the tools collection defining relevant geographical data is a methodological challenge: what is the “position” of a movable object? What would be the point in recording its current position (i.e in a Museum’s reserves) from the point of view of scientific analysis? At this stage we record tuples of values: the position of the commune where an object was created and the commune where it was used.

This at least paves the way for a visual analysis of exchange routes (rather basic though, see Fig. 6). However in many cases we only have an indication on the département of creation and of use, and the visualization then is ineffective. It is plausible anyway that a more significant information would be a simple time+space assessment of presence.

D. The temporal dimension

As mentioned before, temporal data recorded in the edifices collection are “time anchors”. This indication is, as often in the heritage field, dramatically impacted by the multivocality and interrelatedness factors. In many cases the info we harvested is contradictory, and when not its level of independence is hard to state. However this case study also underlined a number of other potential pitfalls: determinacy (numerous edifices are simply not dated), scope (what is actually dated is not clear – presence at time T, time of construction, etc.), accuracy (wordings like “in the middle of the 16th century”), periodization (“after the middle ages”). For each edifice we record as many time anchors as read from the sources, and have defined a “conventional temporal mapping” grid that helps us transferring the verbal indication (“beginning of the 17th century”) into a quantified time slot (“1600 – 1620”) used at visualization time. This solution was at start designed as a makeshift solution, and we are now investigating how an ontology of temporal hints verbalization modalities can help in the analysis and visualization steps.

In both the crafts and farming tools collections what we record are time intervals during which the item is “active”, as well as hints on for instance intermittency patterns. Cross-examining the three collections showed that ultimately one of the challenging issues we need to address is harvesting and reasoning not only on a chronology, on a linear time succession of dates, but on the overlapping of recurrent temporal patterns as implicitly present in expressions such as “every summer”, “every second Sunday of July”, “during the harvesting of the grapes”. Consequently, we complement the data models that are specific to this and that collection by a data model used to collect and document cyclic (or at least non-linear) temporal occurrences (markets, fairs, votive festivals, pilgrimages, etc.).

The data collected acts as a potential bridge between collections, since in a number of cases such an occurrence concerns two or three collections (e.g. a fair where a craft was presented, tools were sold, during a votive festival in relation with an edifice). In parallel, that data acts as a test bench for the formalization and visualization of cyclic temporal occurrences, including in the verbalization modalities.

IV. VISUALISING THE DATA

The general objective in visualizing the data is basically to cross-examine, correlate, and perform reasoning on the temporal, spatial, thematic dimensions, to spot significant patterns with keeping a concern for doubtful info. Over time we have experimented a rather large number of cartographic solutions, and of InfoVis solutions, with for each an ambition to try and answer to specific questions (a tribute to J. Bertin’s

vision) such as *is there a relation of the altitude to the orientation of edifices? Are there more contributors in one commune than in another one?* The inspiration behind these experimentations is at the intersection of geovisualisation, time-oriented data and InfoVis, an intersection personified by the works and legacy of C.J Minard [23]. In this section we present three examples that we believe show the above methodological intersection is worth exploring (even with “poor” data sets in terms of quality and consistency).

A. A map of exchanges

Fig. 6 shows a leaflet-based map [24] produced in order to explore the basic geographical data we have for the farming tools collection, i.e. commune of creation, commune of use. The visualization as such, a simplification of the flow map visualization paradigm, appears quite efficient in highlighting some tendencies, for instance the exchange routes between winter and summer areas concerned with transhumance related activities (SW \leftrightarrow NE). Colors (see legend), are used to represent the amount of objects created (interior circle) and used (exterior circle) in a commune.

Two communes (Arles and Névache) appear as strong “exporters” of tools. It is important however not to jump to conclusions: the data set is too inconsistent to provide a robust analysis ground – the visualization from our point basically acts as a proof-of-concept and shows what could be gained from a larger and better documented collection.



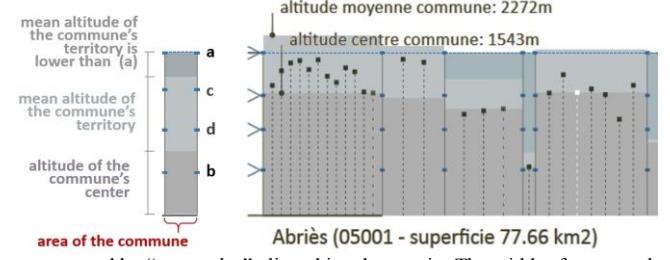
Figure 6. A visual representation of the communes of creation and use – circles represent communes, curves are used when a tuple [creation, use] info is associated to a tool, and the thickness of the curves correspond to the amount of tools. (a - St. Véran, b - Arvieux, c - Névache, d - Arles).

B. Densities and altitudes

The area of concern combines very different landscapes (mountains of the Southern Alps, dry hills, a coastal strip densely populated today, *etc.*). What can be learnt on the spatial distribution of edifices with regards to factors such as landscape, climate, accessibility, and so forth? The following visualisation (Figures 7, 8, 9) helps reading an interaction of parameters: area, various altitudes, density of edifices per commune / per department. Some striking observations emerge, that sometimes obviously deny false beliefs. For instance the “classic” vision of Provence’s hilltop villages that dominate the land below appears scarcely in line with reality. In the alpine valleys climate and accessibility constraints, along with the summering activities, lead to

impressive densities of edifices in territories deemed as non-wealthy. In Fig. 7 the commune of *Abriès* is underlined – almost all edifices are higher than the commune’s center, with a rather regular layering that is typical of mountain territories inhabited for a long time span.

Figure 7. Composition of the graphics: communes of a département are



represented by “rectangles” aligned in a long strip. The width of a rectangle represents the area of the commune. Blue bars on the y axis represent min,max and average altitudes concerning a département : **a** - highest center of a commune inside a département, **b** - lowest center of a commune inside a département, **c** - mean altitude of territories, **d** - mean altitude of centers of communes . In the case of a mountain commune like *Abriès* the mean altitude is unsurprisingly higher than this of its center. Vertical lines topped with a little squares correspond to edifices inside each commune, indicating each edifice’s altitude. For non-localized edifices both lines and squares are white, and aligned with the altitude of their commune’s center.

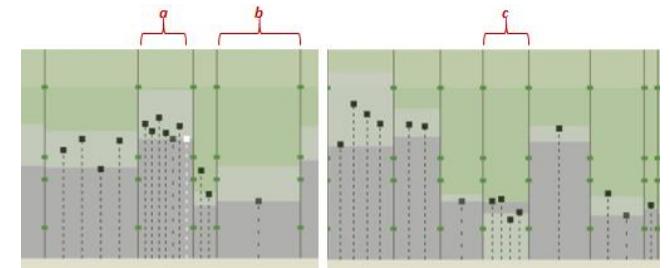


Figure 8. Left, note the dramatic difference in terms of density between **a** (Beauvezer) and **b** (Bevons) – although the latter is lower in altitude, and larger too. Right, **c** (Lurs), exemplifies a “hilltop village” pattern, with the center of the commune higher than its mean altitude.

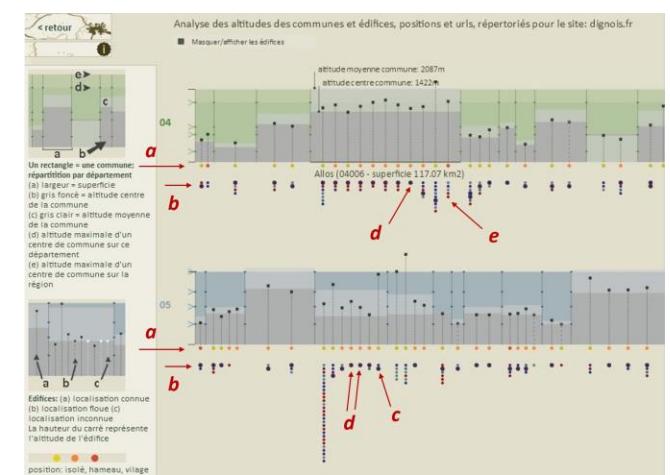


Figure 9. A version of the densities and altitudes visualisation customized for each information provider (produced on the fly, upon selection of a given information provider, here *dignois.fr*). Only those communes for which the information provider does provide e-content are represented.

This version of the visualisation includes indications on the position of the edifice (**a** – isolated vs. hamlet vs. village). It also shows information about e-sources recorded for each edifice (**b** – colours of circles correspond to the type of information providers, see Fig 5). The e-sources produced by the information provider are represented by a larger circle (**c**). The visualisation can be used for instance to spot inside a given commune edifices that the information provider is the only one to document (**d**), or by contrast edifices that the information provider does not document.

C. The “orientation cloud”

Christian churches in Europe are, in theory, oriented edifices at least until the Baroque Period (the apse should face the East). Architectural treaties teach us how middle ages builders managed to apply that rule using kind-of basic gnomons. But was that rule really applied to small chapels? Until when? Does the relief or the altitude impact its application? When the rule was not applied, did builders orient edifices erratically? In this visualization we correlate the orientation with three variables: container (département), date (time anchors), and altitude. Each rectangle corresponds to an edifice, positioned around a “compass”: edifices with an apse facing the north are positioned on the “north” of the compass. The visualization underlines for instance a tendency of older edifices to be better in line with the rule (Fig. 10). It also shows an inclination of later builders for turning the apse towards the north, and 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).

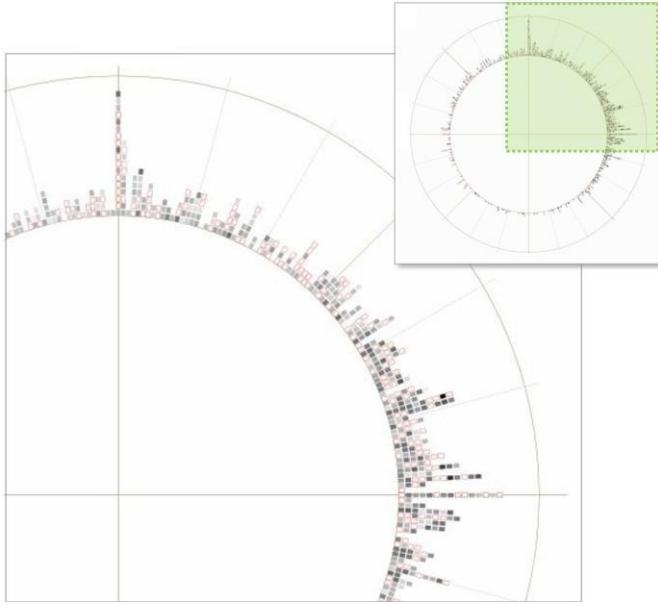


Figure 10. Orientation cloud (partial view showing the most dense part of the “compass”, the North-East sector) – the darker the rectangle, the older it is. Outlined in red: non dated edifices..

The approach is implemented as a web platform combining classic components: an RDBMS (MySQL), a web Server (Apache), a large number of scripts (Perl / PhP) used to produce (on the fly and/or as files depending on the need) textual content : HTML files, JavaScript variables and

arrays, csv “raw data” tables, SVG (visualisations) or geojson data sets. The interaction between those components is operated through JavaScript components developed on purpose. The cartography bases on the leaflet library [24].

V. EVALUATION

On the overall thirteen visualisations have been produced up to now, with three of them “customized” for each Information Provider (Fig. 2, Fig. 9, Fig 11). Their evaluation has been carried out in a twofold manner: an assessment of their *usability* and *understandability* for non-experts (no familiarity with infovis practices nor with minor heritage), and an assessment of added-value for Information Providers themselves, carried out during ad-hoc workshops. During those workshops, we also collected a more general feedback on the research’s fundamental approach, a feedback we will discuss in the final version of the paper if accepted, as the last workshop still needs to be carried out.

A. Usability and understandability

The objective of this first round of evaluation was to get a feedback on the graphic choices (encoding, layout, interaction, support for users), but also on the efficiency of the visualizations, *i.e.* on whether or not they do support *information discovery*. To do so six non-experts were asked to fill in one form per visualisation, with each form structured according to the following set of criteria:

- readability assessment (measuring to which extent graphic elements – colours, shapes, etc.- are easy to identify, and to differentiate from one another);
- problem accuracy (feedback on parameters that users think should be taken into consideration) [25];
- knowledge communication (Evaluation of the capacities of the visualisation to help users understand pieces of knowledge or to help knowledge holders transfer their knowledge [26]);
- reasoning and hypothesis generation (Evaluation of how the visualisation supports hypothesis generation and interactive cross-examination of data [26]);
- adoption and reuse (feedback on the potential dissemination constraints) [26];
- user guidance (measuring to which extent means to guide the users are relevant and efficient) [27];
- legibility (Verification of the lexical characteristics of the information) [28];
- adaptability (capacity of a system to behave contextually and according to the users’ needs and preferences);
- consistency (the way interface design choices (codes, naming, formats, procedures, etc.) are maintained in similar contexts, and are different when applied to different contexts).

We presented to the testers the project as a whole, the data sets and the general relations between pieces of data (e.g. *edifices* located in *communes*, documented by *web sources* that are classified according to *types of info providers*, etc.). The evaluation was then performed by each tester on his own, under no supervision: the visualizations

themselves were not explained, and testers had to analyse them basing on what they were told of the underlying data, and on the legends.

Answers we collected through that evaluation helped us rethink some basic and early choices such as colour palettes, sizes of graphic elements, distances between elements, level of detail of the legends, or the integration of graduated scales in order to help reading quantities. We nevertheless acknowledge the fact that such an evaluation, given the small number of testers, and the absence of a trial and error process, should not be considered as more than a starting point. The fact that testers had no background knowledge or involvement in heritage sciences at large leads to an inconclusive evaluation on factors like “problem accuracy”.

B. Added-value for Information providers

As researchers, we have made choices in the identification and modelling of the information, and then developed a series of visualizations thanks to which we could perform reasoning tasks and formalize an analytic discourse. But, after all, is all that effort of any help to information providers? Since the information providers we have called in are unfamiliar with information visualization solutions, will these solutions be of any use for them? And if they are not, are we really into something that has to do with citizen science?

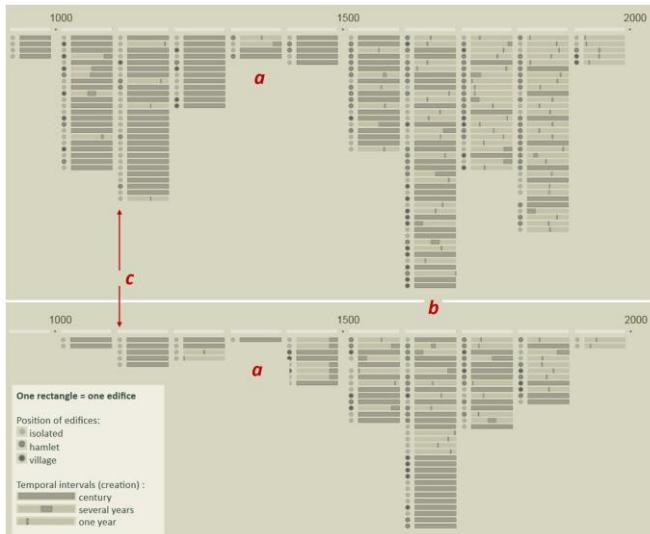


Figure 11. The customised chronology visualisations for two information providers (top: digneis.fr; bottom: montniece.fr). Each rectangle corresponds to one edifice, the diagram is read from left to right, with each column corresponding to occurrences inside a century (partial views of the visualisations, period prior to 900 AC not shown). Bottom left, legend: graphic encoding of parameters “position of the edifice” and “period of creation”. The data sets correspond to individual actors engaged in the study of two strongly contrasting territories inside the region: digneis.fr focuses on one of the the least populated french departements, and one of the poorest; whereas montniece.fr focuses on the hinterland of the Nice - côte d'azur territory, a more densely populated area, relatively recently integrated to the French Republic, and not one of the poorest French departements. Note, nevermind those differences, strong similarities in **a** - a significant rupture in the creation activity during the XIVth century (plague epidemics are a potential, however still unproven, explanatory factor) or in **b** - an intense creation activity during the XVIIth. Also note, in

c, an observation that can be made on both data sets: old edifices are in majority isolated edifices – a common sense assertion (no need to erase an edifice with no neighbours...) here backed up by facts. But beyond similarities between the features of the data sets the visualisations underline significant differences: note for instance the proportions of “prior to 1500” edifices in both cases.

In a paper introducing visual analytics, [29] coined the key expectations that need to be met: *understand and analyse our data, understand and analyse our analyses*. It was therefore important for us to get direct feedback from information providers on the visualisations. To do so we organized a series of workshops during which we privileged open discussions on each of the visualisations, including the customized ones. It turned out that they do provide “food for thinking” services on such aspects as density patterns or information quality patterns, and help spot “errors” (typically contradictory data, misinterpretations of temporal hints, etc.).

But beyond that, and somehow beyond our expectations, the visualisations brought a significant support to workgroup exchanges, for instance on temporal distribution patterns as they emerge from the works of different IPs (Fig. 11). Ultimately nine out of the twelve IPs included in this second round of evaluation expressed the will to get involved in carrying out more such experimentations with us, which is a somehow serendipitous outcome (as well as a fulfilling one).

VI. LIMITATIONS AND CONCLUSION

This research addresses some pending issues at the intersection of minor heritage preservation, citizen contributions, and spatio-temporal data analysis. The paper focuses on three challenging aspects for academics who would want to build on citizen-birthed information sets in order to better document and analyse minor heritage items:

- The heterogeneity of the Information (in terms of scope, of editorial choices, of quality, *etc.*), and behind it the heterogeneity of the Information Providers themselves. We present and discuss the strategy adopted in order to demonstrate potential added-values of such a research on *both sides* (academics as well as information providers), and in order to pinpoint *profiles* of information providers.
- The *factors of imperfection* that are likely to be met when handling such information sets. We first give a global view of the data and information harvested from citizen-birthed e-sources, and then propose an exemplified list of the key factors of imperfection we came across during the research up to now.
- The design and implementation of visual solutions supporting analytical tasks in the specific context of *imperfect* information sets. We present and discuss the learnings of some of the solutions we have developed, solutions that are today available online (*territographie.map.cnrs.fr*).

The case study acts as a test bench helping to investigate data harvesting and visualization challenges. It shows that there is still a significant effort to make in adapting contribution modalities to heterogeneous collections and to the nature of the information we target. It also shows that

investing time on the visualisation step, despite information imperfections, is sound: significant patterns emerge. But these patterns are not assertions: *what they renew is our capacity to question and challenge our own level of knowledge and of understanding* of those collections.

Some clear limitations should be quoted, though, at this stage of our research. Making a sound and grounded assessment of the initiative's added value is arduous since its impact depends on a time taking effort to call contributors in, and to analyse feedbacks. We therefore make no claim that we can present definitive conclusions, but on the overall we believe the experiment shows there is before us a shift in the way academics and collection holders can decode, re-read, augment minor heritage data sets: a shift from "one shot, one collection" protocols to "comparative, cumulative, open science" investigation modalities.

ACKNOWLEDGMENT

This research is funded by the région Provence-Alpes-Côte d'Azur regional authorities, and conducted in co-operation with the Mucem (Musée des Civilisations de l'Europe et de la Méditerranée) - authors thank E. De Laubrie and Y. Padilla.

REFERENCES

- [1] W. Kienreich "Information and knowledge visualisation: an oblique view". *MiaJournal*, vol.0, No.1, 2006.
- [2] D. Keim, J. Kohlhammer, G. Ellis and F. Mansmann (Eds) "Mastering the Information Age. Solving Problems with Visual Analytics". Eurographics Association, 2010, <http://diglib.eg.org> (retrieved 21/04/2018).
- [3] N. Gershon. "Visualization of an imperfect world", in *IEEE Computer Graphics and Applications*, 18(4), pp. 43–45, 1998.
- [4] K. Koszewski "Visualization of Heritage-related Knowledge – Case Study of Graphic Representation of Polish National Inventory of Monuments in Spatial Information Systems", in *Envisioning Architecture: Image, Perception and Communication of Heritage* / Kępczyńska-Walczak Anetta (red.), 2015, Lodz University of Technology, pp. 377-387.
- [5] D. Myers, A. Dalgity, I. Avramides "The Arches heritage inventory and management system: a platform for the heritage field" in *Journal of Cultural Heritage Management and Sustainable Development* Vol. 6 No. 2, 213-224 Emerald Group Publishing.
- [6] P. Le Boeuf, M. Doerr, C.E. Ore, and Stephen Stead, S. (Eds) "Definition of the CIDOC Conceptual Reference Model" version 6.2.3, available at: http://www.cidoc-crm.org/sites/default/files/2017-12-30%23CIDOC%20CRM_v6.2.3_esIP.pdf (accessed April 12, 2018).
- [7] A. J. Jokar, A. Zipf, P. Mooney, & M. Helbich. *OpenStreetMap in GIScience: Experiences, research, and applications*. Cham: Springer (2015).
- [8] M.F. Goodchild. "Citizens as sensors: web 2.0 and the volunteering of geographic information", *GeoJournal*, 2007, vol. 69, pp211-221.
- [9] L. See, P. Mooney, G.M. Foody, L. Bastin et al. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information*, 5 (5) (2016). 55/1-55/23. ISSN 2220-9964
- [10] P. Gautreau; M. Noucher. "Sharing Platform in Digital Geographic Information: Everything It Promise?" *Justice Spat./Spat. Justice*, 10, 2016.
- [11] M. Haklay. "How good is volunteered geographical information? A comparative study of OpenStreetMap & Ordnance Survey datasets". *Environment & Planning B: Planning & Design*, 34, 4, 682-703, 2010.
- [12] S.Spyratos, M. Lutz and F. Pantisano. Characteristics of Citizen-contributed Geographic Information, Huerta, Schade, Granell (Eds): *Connecting a Digital Europe through Location and Place*. Proc. of the AGILE'2014 International Conference on Geographic Information Science, Castellón.
- [13] M. Ridge, M., ed. *Crowdsourcing our cultural heritage*. Ashgate Publishing, Ltd., 2014.
- [14] J. Noordgraaf, A. Bartholomew and A. Eveleigh. "Modeling crowdsourcing for cultural heritage", *Museums and the Web 2014: selected papers from an international conference*, Silver Spring, MD: Museums and the Web LLC, 25-37.
- [15] A. Keinan A. MicroPasts. An Experiment in Crowdsourcing and Crowdfunding Archaeology. *British Archaeology*, N.139, 50-55, 2014.
- [16] D. Freeman, J. Freeman. *Use your head: the inside track on the way we think*. London: John Murray, 2010.
- [17] A. Wiggins and K. Crowston, "From Conservation to Crowdsourcing: A Typology of Citizen Science," *2011 44th Hawaii International Conference on System Sciences (HICSS)*, Kauai, HI, 2011, pp. 1-10.
- [18] W. Aigner, S. Miksch, H. Schumann, C. Tominski, *Visualization of Time-Oriented Data* Springer: Human-Computer Interaction Series, 2011.
- [19] M. Skeels, B. Lee, G. Smith, G. Robertson. "Revealing uncertainty for information visualization", 2010 Macmillan Publishers Ltd. 1473-8716 *Information Visualization* Vol. 9, 1, 70-81, [on-line] <www.palgrave-journals.com/ivs>
- [20] J.Y Blaise, I. Dudek. "Picturing What Others Know: Towards a Dashboard for Interdisciplinarity". In Proc. of the 14th I-Know International Conference 2014 (pp. 15:1–15:8). New York, NY, USA: ACM.
- [21] Thomson J., Hetzler B., MacEachren A., Gahegan M., Pavel M., Typology for Visualizing Uncertainty, In Proc. of the SPIE-VDA 2005: SPIE/IS&T, (Conference on Visualization and Data Analysis), 16-20 January 2005, San Jose, CA USA
- [22] Zuk T., Carpendale S., *Visualization of Uncertainty and Reasoning*, [in] A. Butz et al. (Eds.): SG 2007, LNCS 4569, Springer-Verlag Berlin Heidelberg 2007, pp. 164–177
- [23] Friendly, M. *Visions and Re-visions of Charles Joseph Minard*, *Journal of Educational and Behavioral Statistics*, 2002, Vol. 27, No. 1, 31-51.
- [24] Leaflet open-source JavaScript library for mobile-friendly interactive maps. <http://leafletjs.com/>
- [25] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series, CRC Press, 2014
- [26] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, S. Carpendale. *Empirical Studies in Information Visualization: Seven Scenarios*. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers, 2012, 18 (9), pp.1520–1536.<<https://hal.archives-ouvertes.fr/hal-00932606>>
- [27] S. Beier. *Reading Letters: designing for legibility*, BIS Publishers, 1-190, 2012.
- [28] J. M. Christian Bastien and Dominique L. Scapin *Ergonomic Criteria for the Evaluation of Human-Computer Interfaces* Technical report N° 156 INRIA May 1993 http://www.coccoheads.fr/wp-content/uploads/files/Ergonomic_Criteria.pdf (accessed April 21, 2018).
- [29] D. Keim, G. Andrienko, J.D. Fekete, C. Görg, J. Kohlhammer. "Visual Analytics: Definition, Process and Challenges". *Information Visualization - Human-Centred Issues and Perspectives*, Springer, pp. 154-175, 2008, LNCS.