Cultural Heritage digital data: future and ethics

Filippo Diara[1]

[1]PhD student in Architectural and Landscape Heritage
Architecture and Design department (DAD)
Politecnico di Torino, Castello del Valentino
Viale Pier Andrea Mattioli, 39 - 10125
filippo.diara@polito.it

Abstract. Actual technologies are changing Cultural Heritage research, analysis, conservation and development ways, allowing new innovative surveys. Terrestrial and aerial laser scanner, terrestrial and aerial photogrammetric techniques, GIS (Geographic Information System) and remote sensing techniques are nearly much needed methods in the Cultural Heritage field. These survey techniques produce different kinds of data that needs to be managed in the best way possible. Ethical questions come up about the future of these data. It is nearly necessary dealing with problems like data storage, hardware and software relationship and data redundancy.

Keywords. Accessibility, Cultural Heritage, Digital data, Ethics, Future

1 Introduction to data ethics

Data (and big data) ethics can be considered as a distinct branch of applied philosophy. The focus on big data is concerned with what is being processed, the nature of what is being processed, the findings of analysing the data and who the processing is being done for o by [06].
Technology and computer science are part of the Cultural Heritage research from a long time. All digital data produced is strictly linked to hardware, software and network issues, and all these issues need to be managed for the data sharing, future accessibility and usability. Another important point is linked to the economic aspect, because most of these technologies applied to Cultural Heritage, such as geomatics technologies, are high cost solutions, and this reason bring us to reflect always on ethics.

2 Cultural Heritage actual technologies: metric survey and data management

Metric survey, and then all geomatics techniques, is a necessary step of knowledge in the Cultural Heritage field, because it allows to register metric and geometric information in order to represent, communicate and preserve the surveyed object. This kind of survey, a 3D survey, can be carry out using different technologies, such as LiDAR and photogrammetric approach.

2.1 LiDAR technique

Laser scanner (terrestrial and aerial) is a fundamental method included in LiDAR technique (Light Detection and Ranging or Laser Imaging Detection and Ranging). It is a *range based* technique because it produces electromagnetic signals that later are recorded into the machine in order to derivate metric information [09]. Laser scanner survey represents a versatile solution about big data acquisition, for example a large amount of information of complex architectures. Then the machine has the laser that, with a light beam (radiation beam) formed by photons, provide the necessary metric information to reconstruct objects and environments [02]. Touching materials, radiation beam has different behaviours: it is partly absorbed by the same material and partly reflected and relayed. Absorption is the relation between absorbed energy and incidental energy while reflectivity is the relation between reflected energy and incidental energy. The principal quality of laser scanner is the acquisition speed[1]. Moreover, the single point, touching the intercepted surface, is defined in a reference system where its origin is at the machine heart, with specific x, y, z, coordinates [01]. The ability to manage a large amount of information derived from laser scanner survey is an outstanding problem and the ability to reproduce high quality 3D models depends directly on software and processor power development.

[1] It allows to obtain scanning sections up to 400.000 points per second. The final product of the scanning process is called point cloud or range map.

2.2 Photogrammetric technique

In addition to LiDAR techniques, the photogrammetric method applied to Cultural Heritage has experienced in the last years an enormous growth. Photogrammetry (UAV and terrestrial) is a technique that allows to obtain metric information, and then the geometric reconstruction of objects, starting from images (*image based technique*) and for this reason is based on photographic sensors [08]. Photogrammetry is therefore a method that allows the stereoscopic reconstruction of objects using a couple of pictures or more (taken from different points of view).
Photogrammetry, as well as LiDAR, can be helpful in case of damaged buildings or degraded archaeological objects, because photogrammetric survey (including low-cost surveys) allows to obtain volumetric data useful for future reconstruction or restoration, and it can be also helpful to freeze a particular situation.
The quality of photogrammetric survey depends on the images (pictures quality in terms of resolution), but many subjective and objective factors affect it, like the technical capabilities of the operator, shadows, reflexions, reverb and others issues, unlike the laser scanner that has optimum performance both day and night (ambient light), then without illumination. Innovation in the computer science and photographic field has encouraged the development of technologies applied to photogrammetry, allowing the software development for different aims (low cost and high cost) [03].

2.3 Storage and data management

The storage and data management useful to document a Cultural Heritage asset can be carry out through the utilization of relational and spatial DBMS (Data Base Management System) and others platforms. Moreover, for data management, GIS systems (Geographic Information System) have a fundamental importance the Cultural Heritage field.
GIS systems, initially used only for the military research, are based on storage and managing of geographic information. Briefly are techniques based on the satellites usage to obtain geographic information (geographic coordinate) and graphic information (digital images).
GIS solutions works with layers, in which vector and raster file can be located. Therefore, working like a CAD, GIS software allow to create shapes and lines in order to represent particular conformations located for example in a satellite image or a historic map. This possibility has a great importance in Archaeology, for example to locate archaeological sites and other points of interest in order to compare them in a unique platform.
As far as documentation and management are concerned, Building Information Modelling (BIM) is still in a growth phase and it represents a real challenge for Cultural Heritage assets. BIM, briefly, extends the possibilities of a traditional CAD capability allowing to establish geometric and semantic relations between 3D objects (surveyed as they are) and external data. Indeed, one of the most important capabilities of BIM technology is to combine the geometric information with data of different nature[2].

3 Hardware and software relations in the Cultural Heritage field

Main technologies applied to the Cultural Heritage filed, described briefly before, are affected by computer science relations and issues. From the technical point of view some lines of reasoning are linked to the used instrumentation, in particular hardware related to the data management.

3.1 Storage systems

In these years the technological development was so fast to make, in proportion, a large amount of data storage (often redundant data). Anyway, this growth must be coupled by the same evolution of the storage systems. Over about ten years we moved from the classic storage over magnetic disks supports (hard disk - HDD) to the best performing (greater reading and writing speed) but delicate (their performance can be change owing to temperature and power supply method) flash supports (solid state - SSD).
The reading and writing speed of the storage devices represents only a small aspect of the ethical problem of this kind of hardware. The data storage capability become fundamental in these terms. Now we are living in the gigabyte (GB) e terabyte (TB) era. By having a handful of GB o TB is like to have a lot

[2] Nowadays there are many commercial BIM solutions. However, these tools are thought and developed mostly for architecture design or technical installations. These commercial platforms are in most cases not suitable for all Cultural Heritage projects. The complexity of the Cultural Heritage assets bring specialists to reflect on the need to create a custom-made instrument. A suitable solution can be guaranteed by using open source software, because they allow to operate directly on source code, allowing modifying, improving, and adapting any tool to specific needs.

of available space, but when we talk about specific professional areas of work the storage space is never enough.
In many academic laboratories the presence of hundreds of hard disks (internal and external) and also NAS systems with a lot of TB (like 40-80 TB) it means anyway to have a few available space, because this aspect is always relative according to the researches and the data quantity and in a few years we can lost a large amount of data due to forced clean actions and bad maintenance.
For ten years the big circulation and evolution of the storage systems based on cloud computing has favoured the data conservation methods and data backup but, on the other hand, it has made us reliant on internet. Many people think that cloud systems are something innovative, although we should talk about a return to the past because cloud computing can be compared to the host centring computing, where in both cases the storage and processing steps are not executed by the user [05].

3.2 Hardware and software relationship

A three-dimensional metric survey or any 3D reconstruction produces files and data that depend directly on the machine's hardware, in particular with two components: the main processor (CPU) and the graphic card (GPU), but the random-access memory (RAM) is very important too. Three-dimensional models, more and more complex and having a lot of information, will demand constantly an up-to-date hardware [04]. Graphic card is very important for 3D data visualization: point clouds, wireframes, polygonal meshes and in particular photographic textures making the most of the GPU's graphic power, especially graphic clock and memory bus. As far as GPU's dedicated memory is concerned, if we got 2GB sometimes this amount might not be enough to run smoothly 3D models: 3D model fluidity means the 360° degrees analysis of it, but at a specific speed[3].
CPU is another fundamental component for the processing and computing power and it makes possible the correct operation of the other components, where, as we've seen, the GPU is maybe the principal hardware element for 3D modelling [04]. Finally, CPU and GPU are very important for the creation and analysis of 3D complex models. Anyway, these components have a short duration before they becoming dated, depending on the workload they are subjected.

3.3 Software and file compatibility

Software compatibility, operative systems (OS) compatibility and more in general different kinds of machines (computer) need to be mentioned. Many professional software (for research and data management) are developed following some guidelines, in particular the higher circulation of operative systems. In fact, is evident that the majority of software is fully compatible principally with Windows operative system, barring the others OS and therefore many people (including free and open source OS). It's only recently that are developing professional software (in Cultural Heritage field) also for all, or almost, operative systems [03].
Another kind of compatibility issue is related to software in comparison with others software: Doing a project using different kinds of software we could have a lot of file with any format and extension. The management of a 3D model file format (IFC, XYZ, OBJ, PLY etc.) involves also to work inside many software and OS platforms and it could be dangerous for file compatibility and then for the project stability in general. Thus, should be necessary planning in details the entire project, including software and OS we could use.

3.4 Photography

In the Cultural Heritage field, such as archaeology or historical architecture, photography (that includes also hardware and software systems) has revolutionized the way to produce documentation, analysis and then also the way to teach. Nowadays photography is used also for other aims: photographic survey has become maybe the main way to do research, with which it is possible to investigate archaeological sites or historical buildings, we just think to the photogrammetric method, orthophotos, 3D photographic models and many others.
Moreover, in the photographic field the technological development was ruthless: photographic sensors, more complex lenses, more powerful micro-processor are making dated any photographic survey done a few years earlier; like a survey of an ancient masonry that needs a consolidation or restoration intervention, where the level of detail of pictures is very important for future studies (inside laboratories).

[3] To run smoothly a 3D model means to *turn* it at minimum 30 frames per second and it also means to have monitors with a good refresh rate

4 Ethical issues in Cultural Heritage digital data

When we talk about Cultural Heritage some ethical questions are inevitably involved, especially in some aspects of research and communication.
The ethical question about data storage is fundamental. Analysing the innovative aspect of cloud storage we can anyway deduce that who have data (also personal and sensitive) into cloud systems is not the real owner but the donor. Without internet connection and then without these cloud services we do not own our data anymore because our data are like hostages in the cloud systems owner's hands. Following this reasoning we need also to involve the sphere of education of children and adolescents about the digital data sharing consequences [010].
As far as the hard disk data storage is concerned, what would happen if in a laboratory or in another workplace all storage devices (external and internal) stop working? How many data and years of research could be lost? Is it safe rely only on a specific storage device? How we can choose the best storage devices to ensure long life to our data?
Doing a project using LiDAR or photogrammetric techniques mean to manage also big data. Here is an example of a research: between 2012 and 2013 we worked into a project about metric survey of a courtyard in the Camaldoli monastery (Tuscany, Italy); this work included also the comparison between two kinds of techniques, laser scanner[4] and photogrammetric approach (just for the entrance courtyard). Final data obtained was useful to understand the differences between both techniques.
Anyway, this project was also of difficult managing, because we obtained up to 200 GB of data (as you can see in Table 1). The point is: how many projects like this in laboratories? How many backups? How many storage systems? How many economic resources to continue same analysis in years?

Table 1: Example of research in 2012-2013. Comparison between two metric survey techniques. Object of study: Camaldoli Monastery (Tuscany).

Technique	**Detected part**	**Main steps**	**Time**	**Data size**
Laser scanner survey	entire monastery	- direct geometry acquisition and 3D points generation	- about 2h and 50 minutes	51 GB
Photogrammetric survey	only the entrance courtyard	- photos acquisition - 3D points generation	- 7 hours for photos acquisition; 2 months for 3D points Generation (with dated hardware)	173,61 GB

Laser scanner survey and photogrammetric survey: **224,61 GB**

This reasoning brings us to reflect that we are living times in which every electronic and computer science purchase (especially hardware) outlives its purpose, becoming out of date in few years. The same thought concern also instrumentation and technologies applied to Cultural Heritage: about twenty years ago the specific instrumentation and machines purchase, in order to supply private or academic laboratories, was considered a real investment, because these instruments offered a huge potential both for final results both for the reliability through time[5]. Nowadays things have really changed: economic crisis has allowed, fortunately, more attention and caution as the economic resources usage and finding requires.
An ethical (and nostalgic) question is about dated scientific analyses carried out with pictures. For example, if we want to perform an analysis about buildings archaeology old studies made out with Rollei Metric 6006 (fully operative between the '80s and '90s) or other cameras we should revisit the original negatives and read them with professional photographic scanner (nowadays quite outdated). Now let's say (but not too much) that research has been carried out in Syria or in Jordan, or anyway in a country with a delicate political situation for different aspects, and the historical building is missing or destroyed. What we can do with the original negatives? What does it matter? Other similar questions could be: How much are the actual photographic surveys valid? How long we could use these kinds of data? How much data of these studies is stored in floppy disks only? In these terms everything is relative and to answer similar questions we have to consider any research and studies precarious, insecure and not permanent, also whether it seems good.

[4] Used by Architecture Department of Florence for another research inside the monastery and it was surveyed entirely.
[5] This is another reason why in many universities the financial security of the period guaranteed too easy purchasing opportunities.

Therefore, we need to confine any instrument and technology, also hardware and software, during the time are used and then we need to consider them temporary. It is the same for the final data of a research: how long we could use final data of an architectural project? How long file like DWG, DXF, XYZ, OBJ, PLY, could be used? Could we use these kinds of file in ten years?

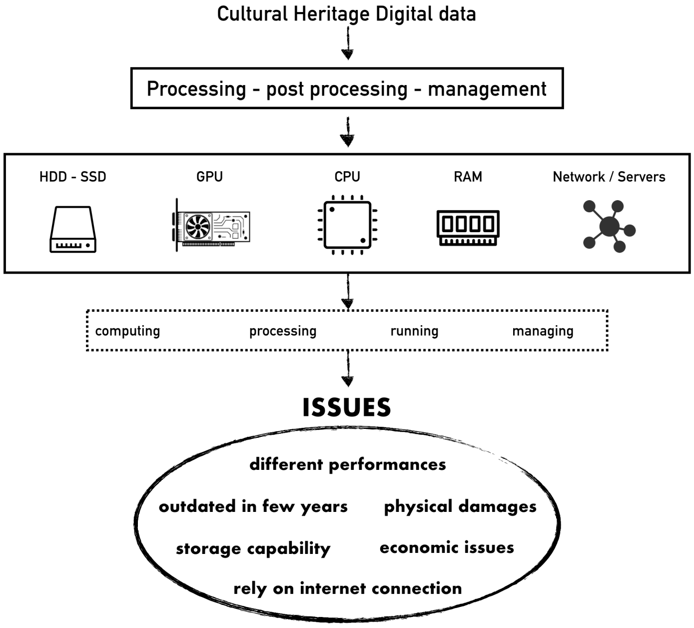

Fig. 1. Schema of hardware complexity managing Cultural Heritage digital data

Moreover, when we produce any Cultural Heritage research involving computer science technologies we create a more or less significant data redundancy. For example, if we are doing an architectural investigation using total stations and laser scanners we need, necessarily, to plan it in detail. Laser scanner, working as an artificial eye, produces a large quantity of data (in this case point clouds) that it requires a discretization and cleaning process, to be done only in laboratory and not at the time of survey. If we want to scan an historical building located into a wooded area or in a dense vegetation area we should, necessarily, discretize point cloud concerning vegetation or woods, because for the purely architectonic survey purposes we do not need it. However, point clouds concerning vegetation is stored by the laser, taking up storage free space, and, if we consider the performance aspect, the laser uses more power and it creates likelihood of machine misfire[6].

A fundamental ethical problem is focused on communication and access to final data. If we produce computerised data, result of digital instruments and technologies, this kind of data are not always available in the same way to everyone. As far as Cultural Heritage field is concerned, a not specialized community will hardly receive scientific technical data. Lately, this problem is an important issue to settle and thanks to new digital technologies (mid-cost, low cost, open source) it is possible to produce and communicate final data in a synthetic and clear way (but not trivial).

Moreover, we need to introduce another fundamental ethical problem: *backup choice*. Spree accumulation of data can bring us towards an information bulimia. We archive everything but we have potentially nothing. Can we back up all our computer files every hours, every days and forever? Can we back up all internet Cultural Heritage data? Obviously not because all our files and internet data (especially these) are dynamic and change every day and every seconds. Every day we have to do important choices about our data (professional, sensitive and other kinds of information).

Let's think about medieval copyists. All of them have made great contribution to our knowledge about the ancient roman poets and writers. We have a lot of poems and papers about Cicerone, Orazio, Tacito but we know nothing about others writers, also prolific and important. Why? The answer is the choice of medieval copyists, the backup choice. We can do the same reasoning for digital data and Cultural Heritage data. Nowadays we have automatic backup with NAS systems or clouds, although it is impossible to save every file we have. What would happen if our computers or hard disks accidentally break down before the backup of data? And what would happen if these storage systems are in a professional laboratory? If we really have to choose between copying a files group or another, what would we will choose? There isn't and can't exist a definitive solution. It is therefore important to find out how we can resolve this ethical problem coming to an agreement to ensure the right future and accessibility of data, choosing the best solution to preserve our important information.

[6] The same reasoning could be applied to a photographic survey. This kind of survey produces a lot of high resolution pictures, that in some cases containing the desired object within photographic framework (*canvas* o *layout*). It is therefore necessary discretize pictures, although outnumbered compared to laser scanner data.

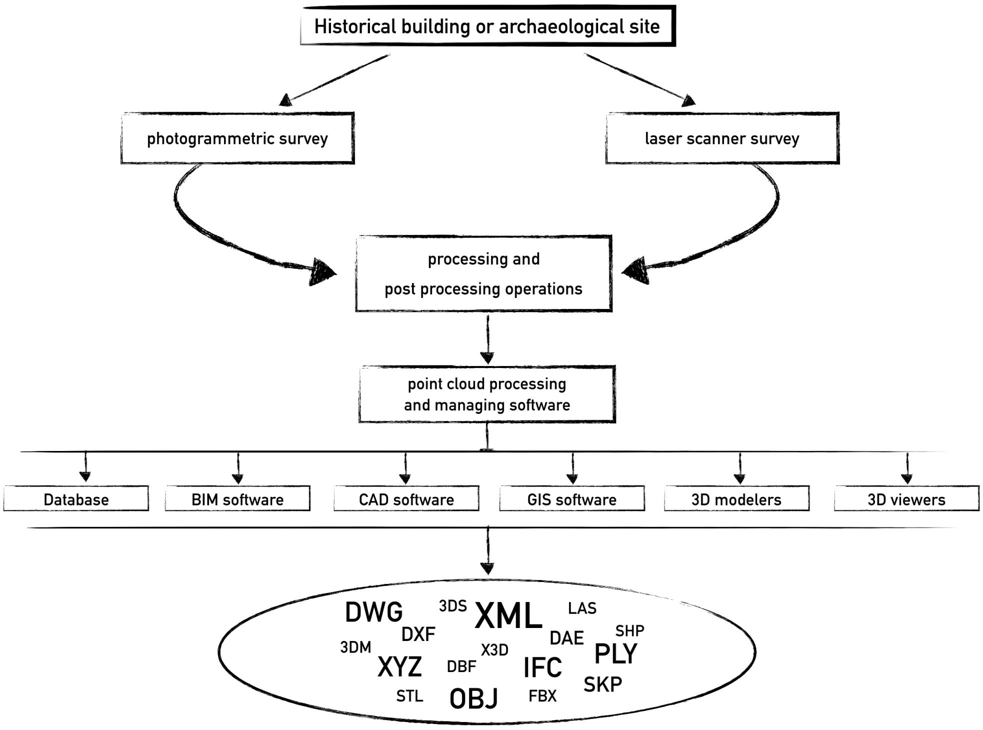

Fig. 2. Schema of software and file format complexity coming up during a digital research in architecture or archaeology

5 The economic issue

After this general panorama, we could say that in order to produce a large amount of digital data we need anyway to use a considerable sum of economic resources. Could Cultural Heritage researchers be considered *privileged*? From cultural point of view certainly of course, but the merely economic aspect is another issue. Could academic laboratories, inside public universities, always afford specific technologies and instruments? The answer is not obvious and it assumes an important ethical problem. On the other hand, who is owner of these instruments sets more accurately in research? Could Cultural Heritage researchers do scientific studies with *low cost* instruments? If it can be done, with equality of results, how these studies can be compared with others carried on with *high cost* technologies? Does it have the same importance? In order to reach final data, at the end of a research, if we use specific computerised instruments (and it has an economic relevance) that bring us to obtain equal results of a study carried on with low cost technologies, behind our methodological choice there is a problem[7].

In the Cultural Heritage field, like archaeology, analyses carried out with *low cost* and FOSS (free and open source software) technologies have often proved to be most reliable, allowing to obtain more complete results than others researches. In these terms, the use of open source and free services could be considered a real challenge, because nowadays the FOSS adoption could be good solution to guarantee the best and complete data usability and accessibility, both for software and file formats. Moreover, FOSS solutions often allow to adapt software to Cultural Heritage needs and not the opposite, thus avoiding methodological stretches.

When we talk about technology and then computer science applied to Cultural Heritage we come up against ethical problems that we could hardly solve them. Nowadays doing research means to use specific technologies and instruments, and it means to do an important economic effort, from public or private entities.

Inevitably, research goes hand-in-hand with economic situations of countries and when we talk about economic situations of countries (funds, investments, etc.) we always run into ethical dilemmas. Therefore, talking in general about Cultural Heritage means (even if into academic areas we don't often realize it) run into sensitive ethical problems linked to economic aspects.

[7] The problem probably is that in the actual society for any objects and purchase there is the wrong conviction that any material object (but also hardware and software) if is more expensive it is more reliable.

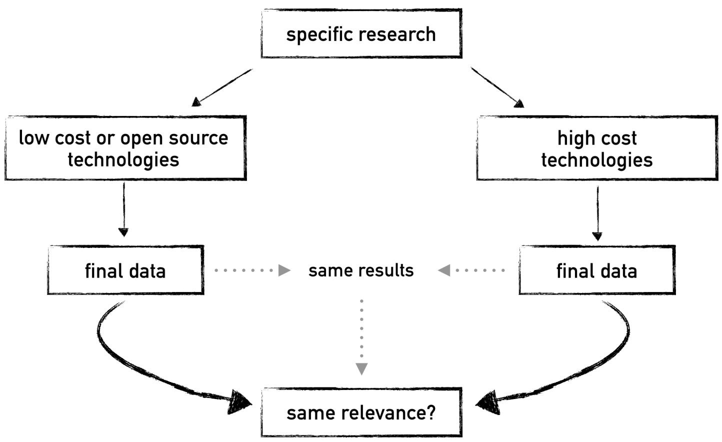

Fig. 3. Economic issue schema of using different budgets technologies

6 Conclusions

This reasoning makes us to suggest that digital data, and in particular Cultural Heritage data are living in an uncertain era, therefore it creates conditions for an ongoing research development. In the Cultural Heritage area, like any area of employment, every day we are given (directly and not) ethical dilemmas for any social aspect and the decision for an equal balance is up to us, in order to secure a good future for digital data and to prevent the abuse of data (especially big data) as a new origin of information and power [010].

All techniques we've seen produce a lot of information and data that need to be archived especially thinking about its accessibility (and when we talk about accessibility should also be included transparency of data) for the future and trying to ensure the future software and format compatibility for next researcher generation. For this reason, is necessary to carry out digital works without all possible new technologies just for fashion but thinking responsible to the aim of a specific research.

At the end we can say as we want responsible innovation (technological innovation) we have to do responsible research, cooperating each other to ensuring a responsive data to ethical issues [07].

References

[01] Bartolucci, D.: Principi di laser scanning 3D, hardware, metodologie applicative, esempi. Dario Flaccovio Editore, Palermo (2009)

[02] Bornaz, L.: I sistemi a scansione laser. La tecnica del laser scanner terrestre. Udine 24-26 novembre (2004)

[03] Diara, F.: Sperimentazioni di nuove tecniche fotogrammetriche di medio e basso costo applicate all'Archeologia dell'Architettura", in *Conservazione e valorizzazione dei siti archeologici. Approcci scientifici e problemi di metodo*, pp. 585-594, 9-12 july 2013

[04] Diara, F.: New software and technologies applied to documentation and communication of Cultural Heritage. Proceedings of "*Conference on Cultural Heritage and New Technologies (CHNT)*" meeting on Vienna (Austria) from 11 to 13 November 2013

[05] Hu, Tung-Hui: A prehistory of the cloud. MIT Press (2015)

[06] O'Leary, D. E.: Ethics for Big Data Analytics. IEEE Intelligent Systems, vol. 31, pp. 81-84, July-Aug. 2016

[07] Metcalf, J., Keller, E.F., Boyd, D.: Perspectives on Big Data, Ethics, and Society. The Council for Big Data, Ethics, and Society, May 23 (2016)

[08] Remondino, F., El Hakim, S.: Image – based 3D modelling: a review. The photogrammetric record (2006)

[09] Remondino, F.: Heritage recording and 3D modeling with photogrammetry and 3D scanning. Remote sensing (2011)

[010] Zwitter, A.: Big Data ethics. Big Data & Society, pp.1-6, July-December (2014)