NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS
Faculty of Business and Management

Tregubova Inessa

MASTER’S THESIS

Developing Data Maturity Prescriptive Indicators as Starting Points of Museums' Big Data Adoption within "Smart Museum" Index elaboration

Degree program: Big data systems

Reviewer
Doctor, Professor
____
First name/Initials/ Last name

Academic Supervisor
Senior teacher
Vladimir V. Opredelenov
First name/Initials/ Last name

Adviser
Doctor, Professor
____
First name/Initials/ Last name

Moscow 2019

Contents

Introduction

Motivation

Nowadays, many countries across the world are closing to the stage of digital transformation that means making public processes and institutions data-driven. The experience of private companies, especially telecom and banks, have proved the positive effect of data on the core processes such as marketing, strategic, resource management and finances. Non-commercial organisations and whole cities, however, start seriously consider the data analytics an essential part of sustainable spatial development as well.

The governments of main world centres as London, Singapore, Barcelona, New York, Dubai have been working on the implementation of data approach into city processes as well as freeing data for the public within the past ten years (Batty, 2018). A concern for data – particularly "Big data" which is directly related to real-streaming and operation of automated functions in the city – is significant. They are moving toward the Smart City vision that implies integration of information and communication technologies, digital devices and physical entities for purposes of planning and control The high-level connectivity between them benefits as citizens as authorities. As a favour, the first group gets easily available public services and higher quality of life. From the second group of view, Smart City innovations improve efficiency, facilitate better planning and help solve urban issues such as air pollution or traffic congestions. The government understands that the data insights taken from various cities departments enable a wider look at cities problems. Gathering data together helps to address issues in the aggregate leading to higher level of local economy and social development and social growth.

In order to accomplish the goal of the creation of a data-rich environment, cities are developing roadmaps specifying for each city department how data will be accessed and exchanged. The most significant illustrations are "Smarter London Together" (Greater London Authority, 2018) and Smart City Lab of Copenhagen. They put forward the bottom-up approach to spreading data employment practice city-wide with particular attention to citizens needs and expectation. Cultural strategy is also a part of proposed initiatives. It includes mapping of cultural venues and events on publicly accessible sources or launching exhibitions in top London's museums which offer visitors to use AR and VR technologies both to enhance their diving into the past and show them the advantages of modern devices. However, despite the important role of culture in the sustainable city development, there is no plan of Smart City development which would explain the place of cultural institutions in the data-sharing process. That is why they mostly stay apart. Unlike transport or the environment programs, cultural one describes either the increase of use of digital devices inside industry or launch of more number of cultural activities thus no taking attention to the role of data in these plans. That leads to the absence of numerical control of cultural institutions performance.

Any strategy development does not exist without a set of metrics for the progress monitoring and, according to international practice, the way toward the Smart City vision is also measured by a performance indicator framework. Such frameworks have many advantages as for the city government as for public services which participate in the strategy implementation. For the government, it is an effective instrument of quick measurement that highlights city strengthens and weaknesses as well as allow to compare their city with another that also helps to understand points to improve or to change. Public services perceive indicators as a recognition of their work done from the one hand and as a motivation to keep developing on the other.

Despite the wideness of such frameworks, they do not consider the cultural sector as a data provider and data consumer. For example, even the European CityKeys project which monitors and compares Smart City solutions across the Europe and which demonstrates the most consistent approach to the assessment of city development(P.Bosch, 2017) does not contain any indicator relates to culture. The framework that includes people health, education, quality of housing and safety but does not mention culture confirms that the cultural sector is skipped in world Smart City vision completely.

In Russia the situation is very similar. Russian indexes created to evaluate cities technological development either do not consider culture at all (Agency of strategic Initiatives,2018) or describe separate innovative devices rather than the whole data strategy (Moscow authorities, 2018).

That is why the idea of Smart Museum Index creation was born. In 2018 CIO of The Pushkin State Museum of Fine arts proposed to elaborate a data management framework which focuses on technological and data maturity. Index includes measurement of all museum processes and roles that enable connectivity and make museums a part of a city-wide network. It has a prescriptive role aiming to help museums formulate their smart vision. Museums plan to propose Index as a part of Smart City Index.

However, the role of Index is not limited by its assessment of museum integration with the external environment. In opposite, its main efforts are concentrated on internal museum capabilities, especially on its work with data. While transport, energetic or housing services started collecting and analysing data several years ago, the discussions about the importance of data analysis in the cultural sphere still continue. As a result, a relatively small amount of museums took the first step to put data-driven decision making into practice. Thereby, the final goal of Index is to drive the development of a data approach in museums.

Looking at opportunities cultural organisations possess in terms of data, first of all, reveals that archives, museums and libraries have huge collections that require automatic monitoring and management. Moreover, since museums and theatres interact with visitors, they should look for ways to personify communications and create a unique experience of engagement for them in order to increase loyalty and have them back regularly. It can only be achieved when applying "Big data" methods that include automatic visitors' data collection and building recommendations systems. Some huge museums as Louvre and British Museum have already started to apply visitors tracking to gather data about visitors' behaviour and preferences that museum staff consequently use to make individual offers and improve the education process.

The special frameworks that called Data Maturity Models are able to help museums to insert KPI of work with data in practice and quickly get a part of Indexes. Such models are spread out in private companies and define the current level of data maturity and direction of progress in terms of dealing with Big Data. In museums, such model would serve as a navigator into the world of data analytics and machine learning. Due to the recommendations they provide to museums, the latter will get aware of the world best practices. This fact will lead to more successful results in museum performance.

Objectives

Following what was said above, we are coming to the necessity of developing an index that would help cultural institution understand how to make data assist them with better fulfilling their cultural and education missions. Example of CityKeys shows that performance measurement framework cannot be both universal and exact for every entity, that is why, this work focuses only on Russian museums, where advantages of data approach are more evident and can be quickly achieved.

The main purpose of this paper is to develop the set of Prescriptive Indicators for Russian museums pursuing the aim to measure their data infrastructure maturity and give recommendations. They will enable museums to be easily integrated into "Smart City" concept and contribute to the enrichment of the world knowledge base. These indicators should become a part of the ongoing Russian museum's initiative "Smart Museum Index" which aims to assess and consequently modernise all current museums' IT-infrastructure in order to make them data-driven. Indicators are designed with attention to both state-of-the-art technologies and institutions financial and legal constraints.

In order to develop the indicators following objectives were formulated:

- Study the existing Indexes created for other institutions
- Study the commercial companies experience of digital transformation and highlight potentially adoptable practices
- Observe the current state of museums' work with data and define weak points
- Define main museum's goals and internal process
- Elaborate system of Prescriptive Indicators and describe how to use them
- Test indicators to evaluate Data Maturity of one museum
- Describe data use cases and its value for cultural sector
- Define project's development prospects

Approbation

As a platform for developing and testing the methodology, the Pushkin State Museum of Fine Arts (the Pushkin Museum further) was chosen. Being one of the museum technological leaders in Russia makes it a good role model for regional museums. Its assessment, therefore, may become a benchmark for other Russian museums. Moreover, the museum is one of the biggest in Russia by the number of hosted objects and visitors number(Pushkin museum website, *2018 year*). With such big values of performance, it has massive potential in data analysis and can become a vivid illustration of positive changes data brings to cultural institutions. Both facts enable effective and complete data indicator framework.

On the other hand, the Pushkin Museum is still an example of an ordinary Russian museum with typical financial issues and legal restrictions that make him an objective reflection of Russian realities. Exploring its activities reveals common museum's goals, principles of operations, main processes and organisational structure that are necessary to define what processes can be improved due to digital transformation and which ones should remain in their current state. Therefore, its experience reveals the common challenges and restrictions that museums have that makes recommendations more relevant.

Structure of thesis

In the first chapter, the motivation for using data by cultural institutions is given by describing museum's roles that can be supported with data. The presentation of three museum's mission: preservation, education and social engagement show museum main objectives and processes that encourage technological development in museum and explain where data comes from. The evidence of better museum fulfilment of its roles due to data collection and analysis is provided.

The second chapter represents a review of the most successful data analytic cases in museums around the world. The aim of the chapter is to define practices that can be easily inherited by Russian museums and as a result transformed in developing indicators. Finally, some gaps and perspectives of museum's data processing are marked that justify the need to turning to the experience of other industries with similar business processes and goals

The third chapter contains a review of the strategies and approaches to data analysis in public and private sectors that are likely to help museums increase process efficiency. Special attention is drawn to a Data Maturity framework which allows to quickly assess and compare levels of data usage among a wide range of organisations and territorial entities.

The fourth chapter describes the development of Big Data maturity indicators framework which matches museum level of proficiency in data with one of five levels of maturity: No data owner, Data aware, Data proficient, Data Savvy, Data Driven. All levels are described in details, but special attention is given to the data-driven state because it is a target.
The fifth chapter introduces the assessment of the Pushkin Museum of Fine Arts big data maturity that is followed by recommendations about further development. Looking at the Pushkin Museum results in wider recommendations to Russian museums, thus making Indicators prescriptions better suit to Russian museum realities
In conclusion, there is a discussion about the perspectives of the digital transformation of the Russian museum industry, in particular, their next steps in progressing with data.

Chapter 1. Museum's missions as motivation for data processing

Global movements towards digitalisation as well as focus on high connectivity between services have an impact on museums development. Significant changes have taken place in the cultural sector across all countries over the last ten years. These changes forced museums not only to look for new means of visitor's involvement or use cutting-edge technologies for collection study but completely rethink the museum place in the changing world. There is a mutual influence of the emerging technologies and new museum's roles because each one gives a reason to exist another one. Innovations allow museum to turn its creative ideas into actions and consequently broaden its set of functions while museums motivate developers to design and create ambitious modern solutions.

In this chapter, there is a literature review of this cooperation which explains museums needs and potential in data analysis. It gives a basis for further research. Provided sources are mostly international since they offer broader and deeper insights from the sector. Examples of Russian experience were included in the second part, which compares Russian and international expertise.

Role of technologies in museum preservation and study missions

First of all, museums are recognised as places for housing world cultural heritage. In terms of this role, the core museum missions are to conservation, catalogue, study and provide access to knowledge (Alexander, 1995)or by other words "all actions aimed at the safeguarding of cultural property for the future" (Clavir, 2002).

Nowadays, museums' work with collection embraces increasingly high number of responsibilities, that are emerging due to technological development. Such requirements as attentive preservation of each object of collection as well as its accessibility to general public and other cultural institutions are met by digital technologies. (Bertacchini & Morando, 2013) Do look at each function separately.

The conservation is a very native museum function because deterioration of collection is inevitable processes that however can be slowed due to climate control activities (Museums & Galleries Queensland, 2014). Museum climatology is dedicated to answer the questions about right air conditions that museum should ensure to best preserve in-house objects. Also it seeks to methods and instrument that can maintain such conditions constantly. The museum microclimate is measured by following indicators:

- Light levels
- Temperature and relative humidity
- Air streams

- Air pollution
- Pests
- temperature and humidity on walls and windows
- noise and vibration

The preservation quality of an environment is best judged in terms of relative risks and benefits to the collection in the space. According to Russian and international authors (Duin (2014), Madsen (2007), Oganesova (2001)) in order to set and maintain optimal conditions museum should complete following tasks:

- defining target values of microclimate features and storage standards
- selection of necessary measurement facilities and methods
- arrangement of data processing and results' assessing
- tuning museum microclimate

However, the storing an object in funds differs from exhibiting it in halls where they are damaged by human intervention, changeable air conditions caused by opening and closing doors or windows. Moreover, being in funds makes it easier for curators to check its quality regularly and prevent possible cracks in advance, while an object that holds for public interests is supposed to be always surrounded by people and cannot be monitored so well.

The additional problems marked in reports of Paul Duin (2014) and Rebekah Butler(2014) are related first to diversity of reasons (chemical, mechanical and biological) why decay happens and second to different reactions of different objects to the same conditions, thus ones that bring benefits for one decay mechanism may bring increased risk with another. Also temperature conditions come particular challenge for exhibition curators because they should consider visitors needs as well.

Nowadays problems of measuring and monitoring are address with mechanical systems which solve them with minimum costs and maximum value. Most museums install (Heating, Ventilation and Air-conditioning (HVAC) systems that are designed to regulate and monitor indoor climate (Madsen, 2007). They usually consist of five components: mixed air, dx cooling coil, heating coil, cooling coil and steam humidification. They are run by the building management system (BMS) that is connected to sensors in the exhibition space and sensors in the air-handling units (AHU). Depending on the outside and return air conditions, as well as on what is the desired input condition, the BMS decides on higher energy consumption due to the greater pressure loss that must be compensated by the intake fan.

However, according to Edgar Neuhaus (2012) HVAC-systems has two risks. Firstly, an occasional HVAC malfunction leads to sharp changes in RH and temperature, especially during winter seasons that negatively impacts preserved heritage and cause its disruption. –For example in 2009 Ancient Art Museum in Brussels in 2009 experienced serious troubles because of the bag in HVAC system that costed 1 million Euros to museum. The unexpected RH drop that was only discovered in 55 days caused the damage of high layer of the stored painting.
Secondly, these systems are not only expensive but also consume lots of energy that also make high costs for a museum. Which is why, when sensors are inefficiently distributed across the building, or some of them are broken, it results in increased costs which shorten museum's already lean budget.

Both these risks can be reduced by the application of automatic methods of data analysis that would regularly monitor data collected from sensors and define or even prevent defects as in stored artefacts as in system components themselves. The detailed opportunity data analysis provides to museum will be discussed in the next chapter.

Climate control is not the only activity that enables museum collection safety - in literature, the preservation role of digitizing collections is estimated very high (Gray, Jo Coones, & Viola, (2017), Anderson (1999), WALKER & TALLON (2009)). Among benefits it brings to the society authors firstly tell about their contribution to documentation and presentation of artefacts that has been collected for thousands of years but now are endangered. Online representation of the cultural heritage means access not only to remotely located objects but even to which of them that are hardly achieved or do not exist anymore because of an accident or human intervention.

The famous example of how digital collection reproduced the real collection that physically disappeared is a case of the National Museum of Brazil in Rio de Janeiro. In September 2016 almost 90 percent of its precious collections were destroyed by the fire and most of specimens and artefacts were irreplaceable. A recently launched Google Arts & Culture project saved them because the institution had recently started to create a digital version of their collections. Google captured museum's interior through "high-resolution photography, photogrammetry, 3D laser scanning, and virtual and augmented reality," writes Chance Coughenour, program manager of Google Arts & Culture, in a blog post. Due to Google work now collection can be seen virtually at the same state as it was before fire.

Furthermore, the accessibility of collection online enables curators to explore in-housed fragile artefacts without touching them. This also decreases the possibility of its destroying and serves for preservation.

Thirdly, being a part of the collection management means a monitoring of changes to which objects are subjected such as replacement or restoration. The full cycle of curation consists of four steps: registration, preservation, storage and exhibiting or research (Robert 1995). Each step should be well documented in order to understand an object's current state and operatively react to negative changes that happen with it. Also, it helps curators to know museum resources better and simplify the preparation to a future exhibition in terms of a topic and exhibits selection.

There are many means of digitization that depend on specimen type as well as intended use of digital copy. Mostly museums choose the range of applied methods with regard to its financial abilities. While one museums are able to commission special companies that deal with 3-D printing or virtual reality, other ones have to use their own human resources and equipment that usually include scanners and cameras. The wide-spread methods are laser scanning, computed tomography, multispectral imaging, 3d image-based model creation techniques and scanning electron microscopy.

Collection management system serves to automate the work with digital objects. While historic ones only catalogued objects, modern ones help museum staff to facilitate preservation function by its fully meaning as it is described above. Its main functionality contains management of objects, management of metadata and reports generation (Canadian Heritage Information Network, 2012). First includes functions that are responsible for object entry process, account for objects, inventory control process, object location and status control, publication and conservation history keeping, cataloguing process.

Obviously, digitation and, particularly collection management systems, generate huge amount of data starting from digital images of objects to their metadata to detailed script of its life. Potentially all this data can improve collection management and make objects issues and threats more transparent and available for curators. The questions about how this data should be gathered in order to benefit museum the best will be described in the next chapter.

Since this type of systems quite popular and their functionalities are similar, providing one example will describe how all of them work. In Russia the most popular system is “KAMIS” that serves to as big museums such as Hermitage and Russian museum in Saint-Petersburg as small museums, for example ethnographic museums of Khanty-Mansiysk.

The system is made up of seven modules: "Accounting operations"; "Scientific and funds work", "Statistics", “Multimedia objects storage”, “Integration with Goscatalog”(new), “Digital copies of accounting books”, “Audit of image download” (КАМИС, 2018). The architecture of the system consists of DBMS (ORACLE), server of applications and thin client (browser). As for working with data, the system does not support any comprehensive analytics but offers conventional statistic reports that cover museums needs while fulfilling preservation function.

The most perspective solutions based on the KAMIS are web-portals of museum of Tatarstan, Sakhalin oblast and HMAO that aggregate information from all small regional museums in one place, thus facilitating preservation of regional cultural heritage and increasing public awareness. That is leading to the next museum role that is a part of the cultural network.

Digitization of collection indeed broadens museum capacity in term of not only collection management but also international networking (Bautista, 2013). Going to online encourages museums to communicate with each other. When all world seeks to be connected museums cannot stay aside. The worldwide access to collections serves to maintain robust relations between institutions and boost global researches that involves efforts of scientists and curators around the world. Gray, Jo Coones, & Ruhse in their book believe interdisciplinary methods and intercultural exchange to be necessary conditions of museum’s adaptation to ongoing changes.

The real exchange can happen only when museums share data and make it linked and automatically readable. Nowadays, when the existence of everything is assessed by its findability in google, publishing data on the website or having it in internal museums systems is not enough, especially for researchers. They are much more comfortable with platforms that aggregate data of different institutions and provide it in logically connected structure than with exploring online archives of each museum separately.

The international example of such platform is Europeana project that serves as a Europe digital library, museum, gallery and archive (Europeana Pro, 2015). The project’s goal is to make European cultural heritage to be “viewed, shared, used and reused wherever and whenever possible”. It is multilingual platform that open access to European collections of books, photos, paintings, television broadcasts to name a few, that are published by cultural institutions which own them. Project makes reuse of data very convenient and motivate to create scientific and educational projects, for example, audio guides. However, making connections in data also means linking professionals that explore this data as well as motivate them work together.

Building such interconnections requires museums to comprehensively approach to structuring their data and releasing it in format suitable for sharing. Therefore, museum should learn how to manage its data smartly and enhance their competence in data processing.

Role of technologies in museum education and inspiration missions

Today museums become the institutions that put interactions with people first. Their social function prevails over preservation work. Certainly, education is historically a principal one in museum work, even during the Napoleon time museum were seen as places to "create citizens" (Alexander, 1995). However, with time, this objective has become more complex, so museum social function has broadened. Not only education but training and life-long learning; not only knowledge but also creativity (OECD/ICOM, 2018). Moreover, some authors go further by naming museums social influencers that shape minds of whole communities. Today, the world's museums perform as agents of well-being and as vehicles for social change. On display there is a growing belief among practitioners, policymakers, and the public alike in the power of museums to inspire hope and healing, improve lives, and better the world. Most of modern studies admit that at their very core, museums today are institutions of social service (Silverman (2010), Suzić (2016), Kotler(2008), K.Jenny (2011), Y. French S. Runyard (2011). Such attitude to museums place in the modern world requires them to attentively build communications with its audience. It is essential to shape a positive image of museum during providing engagement with its collection and public activities. It is the only way for museums to survive in the competitive environment.

Fulfilling the social role means for museum a controlling of two main indicators: the wideness of their audience and the deepness of visitor's engagement (Silverman, 2010). The accomplishment of high values of these indicators means that museum succeeds with becoming a part of people life, that eventually guarantees museum flourishing and fulfilment. However, establishing a trust relationship with society is not an easy task. In contrast, it demands detailed strategy and additional financial, human resources and last time technological resources. Before turning to particular approaches a museum use to achieve high results, first, do have a look at what these indicators imply.

The wideness of audience includes the ambition of museum to attract the maximum number of people without regard to their age, level of education or social status. Recent research of ongoing changes in Russian museums showed that their heads are sure that museums do not have a target audience and work for broad public (П. О. Васильева, 2018). The term wideness is very close to accessibility that means no physical no mental specificities, as well as lack of background, do not prevent a person to visit a museum. Only when all exhibits are explained simply, and space is designed to meet people different needs museum can be called public.

It also should be noticed that wideness of the audience directly influences museum financial success, thus providing evidence of the museum's efficiency for the government. Museums mostly are public institutions which depend on governmental subsidies, which is why they need to prove value of granted money. So, as most visitors pay for the admission (excluding museums with free-entrance) and usually also leave money in gift-shop or cafeteria, each new one brings additional money for museum. Major curator of Jewish Museum and Tolerance Centre in Moscow told in a discussion that their investors measured museum success with regard to number of tickets they sold rather than counted the total price (Strelka institute, 2016). That means attendance criteria is a one of success.

However, Nicolas Cullinan from National Portrait Gallery of London inclines to consider also other criteria such as originality and the historic value of content when make a decision about a topic of next exhibition (2019). He urges to be cautious and do not judge museum only by attendance.

The deepness of visitor's engagement is an essential and comprehensive indicator. While first indicator measures who well museum attracts a potential audience, this one defines the quality of interaction with visitors inside museum. According to Falk and Dierking every particular experience is very individual and depends on such factors as physical, personal and social (2012). Physical ones stand out for what surrounds a visitor in museums. What exhibits and objects he come across forms his impression about the visit. Personal factors include all experiences and memories that belong to a particular visitor and that are likely to reflect important relationships in his life. Social factors include the companions, if any, with whom one visits the museum that a person brings with himself. Silverman adds another one as current visitors that person meets during a museum experience. Museums should take into consideration all of these four factors in order to build successful communication with a person.

The other facet of the deepness is how long visitors stay in touch with museum and how likely he wants to repeat a visit. The opportunity to keep "talking" with a visitor after he leaves the building or maybe even a country gives a chance to enhance the engagement and retain visitors. That is why, time in touch is also a significant indicator that is easily converted into museum financial success and recognition.

In order to increase communication deepness museums adapt new methods and technologies to their routine even if they are not pertinent to cultural sector. Since the wealth of collections is no more supposed sufficient to satisfy the interest of different audience segments, museums urgently need to create unique and impressive experience for a visitor. Technological innovations, particularly services which would generate satisfaction and positive outcomes for museum visitors can be a solution. The particular focus on attraction of young audience can be an additional advantage (Carvalho Ana, 2018), as they are used to getting information in real-time and in interactive way.

The most popular way of technological adoption is digitizing the “conversation” between the institution and visitor that can be achieved by using mediation devices such as audio guides, smartphones, touch-screen tablets, VR and AR. They all exist to augment museum’s objects and make a visit more informative and eventually longer. We also should consider that each device differently impacts a visitor behavior and his emotions (Jarrier Elodie, 2012). Looking closer their main features are following:

- audio guides have a learning function by giving the additional information about the objects (Deshayes, 2004) Moreover it enables the enhancement of the museum experience for everyone, not only for children and families, general public or foreigners, but also for the blind and visually-impaired or people with intellectual disabilities.(Cláudia, 2012)
- online videos stimulate imagination work and the experience of memory (Tussyadiah and Fesenmaier, 2009)
- touch-screen tablets entertains visitors and especially benefits families with children by making the experience more sociable and enjoyable (Gagnebien et al., 2011)
- mobile applications form a knowledge background that prepares visitor and makes him feel more confident

The subsidiary technological solutions such as WI-FI and video surveillance that indirectly impact visitor’s impression also should be mentioned here. Free access to the internet sometimes connects to the quality of the use of the mediation devices, especially, when availability of some functions depends of the speed of internet connection. Although video surveillance usually is not a part of attendance analytics, it also is very important part of museum infrastructure as it guarantees visitors safety.

Audioguide is the most popular device that now exists in almost every museum in the world. However different museums equip them with different functionality as well as expect different outcomes from its usage. It can operate as a separate device or as mobile application. Being as an ordinary service does not stop some museums from create a unique and memorable visitors experience with audiogides. While their classical function is to tell about surrounding objects, there are advanced versions which lead a visitor throughout museum by colliding him with the objects fitted to his particular interests. As a result, the outputs of second ones in marketing perspective will be more significant and allow to draw person attention and make him desire to come back. However, more quality costs more. The smartness of the instrument is a consequence of audience analysis that helps to reveal the common patterns of behaviour and main segments among visitors that are consider during designing a guide to make content better meet visitors interests. The importance of the use of data will be discussed in details later.

The British museum located in London offers for its attendants such sort of “smart” audio guide. With the annual attendance of over 6 mln people audio guide is designed to meet individual needs of each visitor. Being the only product available in visitors’ native language, it mainly targets on international tourists (Bickham, 2016). The guide proposes several itineraries through museum big collection that vary in terms of content for example exhibits’ country of origin or epoch or visitor’s personal factors as previous experience with this museum, time available and visit’s objective. Apart from the information about exhibits, it helps visitor to navigate in numerous halls by marking his location on the plan and instructing him about the shortest way to get the next object. Also it gives additional but still useful knowledge about cafes on his route, the presence of literature related to watched collection in a gift-shop and what is the most pleasurable time and topic of next free tours that are held every day in the museum. Finally, the devices records all exhibits where visitor stopped to listen to the audio and allows to send this statistic to a visitor by email supplemented by the object’s photos and brief description (Pic.1).

— Pic.1 Screen of email from the British Museum with summary of visit

From the visitor’s point of view being guided by a such way produces an impression of intimate connection that has been created between visitor and museum curators during the visit. Such personal attitude encourages him to continue the diving into world history with the British museum next time.

In terms of social role fulfilment, audio guide encompasses every facet of relationships that can be established between human-being and museum. It influences three of four factors of visitor's engagement mentioned above that allows them to achieve high results in visitor's loyalty. Furthermore, the tool builds a bridge between the physical visit and online activities as a virtual exploration of the collection on the website thereby keeping the museum in touch with its audience far longer than only one visit. This fact leads us to another advantage of the technology such as e-marketing that includes full-scale marketing campaign deployed online. It addresses both people who once have visited museum and the potential audience.

In attempt to combine the educational purpose with the capability to involve their visitors through emotions museums turn to more sophisticated than audio guides technologies such as Virtual Reality (VR) and Augmented Reality (AR) (Barbieri, Bruno, & Muzzupappa, 2018). Their popularity has been growing in the last decades due to their capacity to intensify the human senses and bring collections into life. (Coates, 2018). Many museums around the globe have already valued its potential. VR and AR are being used to create museum tours, make exhibits interactive, and to bring scenes to life.

From a content point of view virtual museums can be a one of next three types (Anton M., 2018):

- digital representation of a real museum
- virtual tours around a real exhibition
- websites where user can remotely examine collection in details

The creating of virtual experience brings lots of advantages both for visitors and museum itself (Din Herminia 2007, Capuano N. 2016, Sun 2018, Seadle M.2005, M. Anton, V. Kuchelmeister) . From the visitors' point of view, it means the next level of engagement because it embraces also tangible and visual senses that makes a history exploration more thrilling process. Coming from gaming industry, it broadens museum audience and attracts younger visitors to explore the history hidden in museum. It captivates them and motivates to know more about surrounding objects.

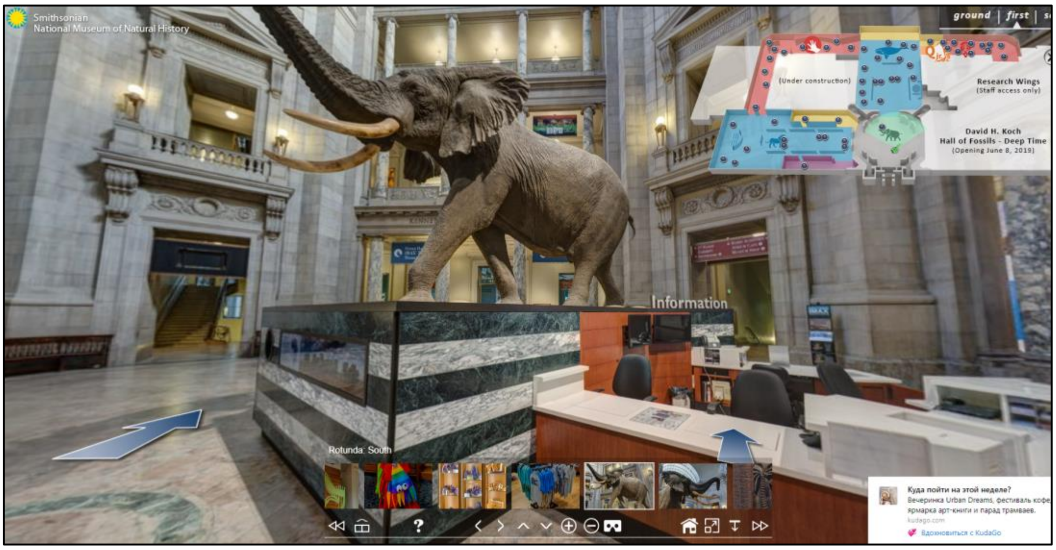
From the museum point of view, it can help curators to put objects in context without addressing complex and expensive interior details. Therefore, VR and AR are seen as a way to decrease the costs of designing and managing the exhibitions. Another advantage of creation 3-d models of the exhibits is their safety because of the opportunity to share the whole image of the object and information about it online. Firstly, it decreases the danger to be occasionally damaged by surrounding crowd and secondly, it gives a time to restore a real artefact carefully supposing it could be temporarily replaced by a virtual copy.

Among the most popular technologies now are the PDA-based virtual museum, portable project-supported museum, head mounted displays (VR –goggles), web-paged integrated online museum. They all make museums “more portable and immersive” (Styliani S, 2010)

Whatever type of technology would be used by museum VR or AR, they all allow to track visitors. Monitoring their behaviour allows curators to understand visitors perception of exhibits around and, as a result, recognise strength and weaknesses of their location. Curators, therefore, find out the points to improve in the exhibition in order to get greater respond from visitor. When such analysis is undertaken at the outset of new exhibition, it decreases designing and managing costs.

In Russia over last two years VR excursions was presented by many museums, including The State Tretyakov Gallery and Hermitage. They provided fully-immersive experience that transports a viewer into the history making him a witness and even a participant of events that complement the physical exhibition. The Tretyakov gallery used the technology to recreate the furnishing of studios of founders of Russian avant-garde Natalya Goncharova and Kazemir Malevich. Within the tour a visitor observes the atmosphere of their work places, looks the creation process held behind masterpieces and finally can draw painting by himself. The main goal of museum was to attract people to the less popular building of The New Tretyakov Gallery and introduce to them Russian art of the 20th century. In Hermitage they created a virtual excursion with famous Russian actor Konstantin Khabensky who guide a visitor through inaccessible for visitor halls of the museum simultaneously diving him into the museum history. Due to this project, visitors can witness famous historic events that took place there. The main particularity of the project is his length as it takes 18 minutes that is much longer than usual VR-tours. Both projects took lots of time and investments to be made however it seems now they have achieved their goals.

Among international unusual cases are National Museum of Natural History of Washington and MOMA of New York First one in the March 2019 launched an immersive experience that let visitor meet a dinosaur T.rex and now everything about his life. The fabulous part of this tour is an opportunity to construct a T.rex bone by bone in collaboration with other participants. Another example is a tour of Metropolitan Art museum available online for everyone. The Metropolitan Museum of Art has a dedicated VR tour project called Met 360º which takes the form of a series of 360 videos from within the museum. The clips are hosted on YouTube and each is accompanied by a short description of the content. The series allows viewers to explore The Met as never before through unique camera angles and the removal of protective covers on certain exhibits.



Pic. 2 Screen from National Museum of Natural history of Washington VR system

All technologies described above serve to help museum educate and inspire audience. However, museum audience is very diverse as we can mention curators are aimed at working with broaden public rather than narrow segment. Ketler wrote “museums serve diverse audiences, multiple constituencies, and visitors who have disparate interests, intentions, and expectations”. The desire to embraces as many people as possible makes museum work with all these particularities. The only way to implement it is to build a strategy based on the audience data insights that requires additional efforts in data processing and analytics.

Role of technologies in museum social function

The mediation devices primarily serve to enhance the visitor’s engagement with focus on education. At the same time marketing technologies such as social networks, emailing newsletter subscription help museum fulfil the role of the public place that enables people social inclusion. It is a relatively fresh representation of museum in the literature that has been being discussed since 2010. All public reports of international cultural institutions say that museums are increasingly engaging in urban life by influencing its citizens and creating communities.

In the report “Culture and local development: maximising the impact” ICOM confirms that today museums contribute to social and economic cities development as “they generate knowledge for and about society, are a place for social interaction and dialogue, and a source of creativity and innovation for the local economy”(2018). Museum activities aim at developing creativity and engaging in the creative process people from different social backgrounds. The view of museum of the future in concerned articles usually represents an open and inclusive public venue in the city where people get together to enter into contact with history, art and finally with each other. Authors believe museums can address real social issues and help people to understand themselves and find their place in life (Jasper, 2016). The importance of digital instruments and especially those that establish relations with audience online is fundamental (Parry, 2013) as they allow museum to constantly stay a part of person's life and indirectly shape his behaviour.

From the end of the last century experts refer to high potential of marketing instruments and digital services in the museum field (McLean, 1994; Gilmore and Rentschler (2002); Kotler et al. (2008); Altınbaşak and Yalçın (2010); Ylva French, (2011), Sandell R., Janes R. (2007)). It has already more than 30 years ago when DiMaggio marked that "although most art organisations are non-profit institutions, they are not non-market institutions"(1985). Kotler affirms that museums need marketing because it enables to provide maximum value for the price of visiting (1998). Since museums have to compete on leisure-time market with other cultural and entertaining institutions, such as cinema or shopping malls. it became very important to offer an experience that overwhelms expectations and is unlikely to be obtainable at other leisure venues. (Ershova, 2017)

In the book of Ylva French (2011) authors note that marketing and PR functions in museum should consider IT as an essential part of PR and marketing functions, because it allows to take advantage of social media development. Michelle Loh (2009) claims that "the uses E-marketing and communications have endowed the modern-day visitor and the museum with tangible and intangible benefits". Among recommended instruments in the literature of the start of 21th century e-mailing, newsletter subscription, news and booking facilities on the website are named as the most efficient (Ruth Rentschler (2009) ,Ylva French (2011)). However social networks and blogging are considered the most productive due to the opportunity of constant communication with audience they provide (Hsiang-Yi L. (2008), Capriotti P. (2018), Bautista S. (2013))

E-mails is a type of communication that embraces the most loyal audience because only people who confirmed their interest in museum news receive letters. Direct mailing is an excellent way of promoting potentially interesting news with low costs that, as a result, makes follower feel involved in museum activities (Mclean, 2012). The only problem of e-mailing is a start point, when people subscribe. The subscribing is possible only on online platform or what is less often on museum's electronic device. This means a necessity to propose another technology to a user first. Here "Museum Friends" program is a possible solution. Its members get a free access to permanent and temporal exhibitions as well as can join private events about which they get know due to the letter subscription. Therefore, on the one hand, program increases the volume of audience that constantly stay in touch with museum and on the other hand, its own members benefit from personal email notifications.

The email targeting is another matter of concern. To be readable, the content of letter should respond a person particular interests that makes museum first produce several versions of one letter and second to define the particular recipient of each of them. Quality of personification in communications significantly increases the response rate. Without special data infrastructure and analytics methods this problem could not be solved well that again leads us to the importance of data analysis for museums.

Email newsletter is introduced now in every museum of the world and all experiences are similar. With a frequency once or twice in two weeks marketing team plans the content and send letters to subscribers about actual events. The difference exists only in additional objectives of emailing. While the most of museums aim only at sharing the news and other relevant content, there are several who perform audience research before sending letter. Such example is the London Transport Museum which in order to improve information about its activities first created a portrait of their audience. However, the cases of personified newsletters were not found, so it is a field to improve for all museum industry.

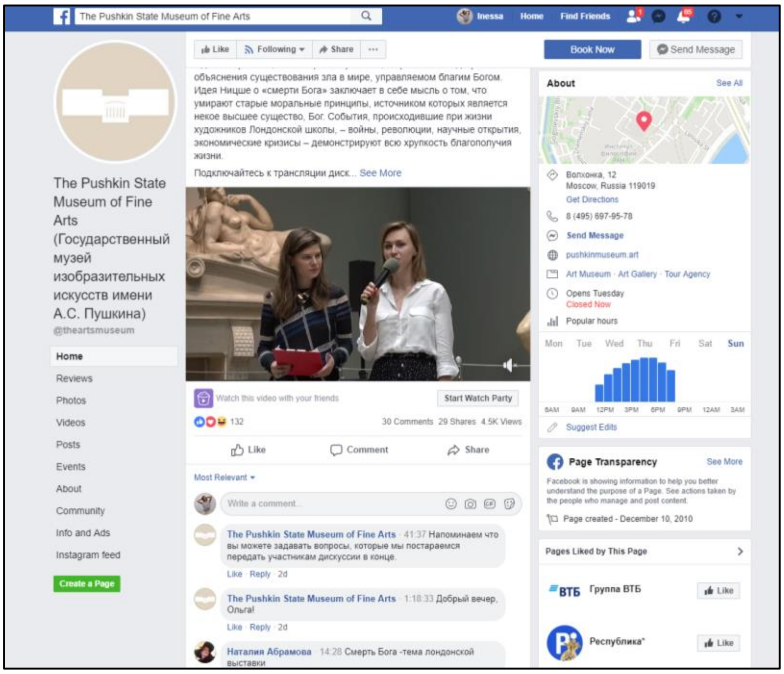
The modern and very effective communication instrument in terms of both count of reached people and quality of communications is social networking. Social networks increase the number of meeting points with a museum, making conversation with museum go beyond its walls and as a result more impactful. (Capriotti P, 2018). Unlike mentioned above instruments social networks do not imply direct advertisement of ongoing exhibitions or souvenirs realization. Instead, they create more participatory and pleasurable ways of contact with public through open discussion and educational content. That is to say, as Holdgaard (2011) notes, through social networks online communication of the museums can be much more active, participatory and dialogic than their traditional communication.

Foreign studies (Sweetser, Lariscy, 2008; Kim, Hoon-Sung 2014; Zafiropoulos 2015; Capriotti 2018) usually give preferences to Facebook and Twitter while in Russia they choose Facebook, Instagram and Vkontakte. It is explained by the outreach of these social networks in Russia and abroad. Another ones as Youtube, Foursquare, Flickr are also mentioned.

In foreign sources Facebook facilitates are presented as helpful for museums to engage in dialogue with their publics being both active listeners and storytellers. Authors affirms flexibility and convenience of the instrument to be a platform of dialog between the organisation and its public. However, according to the study of Capriotti, Losada-Diaz (2018) where they analysed pages of the top 100 art museums worldwide only 4% of posts contain the phrases that provoke sharing opinion or actions of users while other 96% have informational content and usually intend to motivate and prepare users for the physical visit. It shows that museums mostly have a very limited interest in keeping vivid but virtual communications with their public. It can decrease the number of loyal people who stay engaged with museum life after leaving the museum building. In the same research authors found out that almost 100% of posts include text and 10% lesser amount have a picture or photograph. In the contrary, audiovisual resources do not exist almost, their part is less than 3%

The example of the account with "right" posts that urge user to participate is an account of Brooklyn museum (Zafiropoulos, Vrana, & Antoniadis, 2015) They utilize the social network to reach out to young audiences and provide interactive learning tools. Museum actively involves its followers into discussions by creating groups, posting responses to comments, publishing a content that is open for criticism.

Among Russian museums the most followed Facebook page belongs to The State Pushkin Museum of Fine Arts. Apart from news and educational information about collection they also publish broadcasts of lections and scripts of art events taking place in the museum with the aim to invite user for discussion. This leads to many comments from followers and as a results the popularity of the museum page.



Pic.3 Screen from Facebook page of

the Pushkin Museum of Fine Arts of Moscow

Twitter is not a popular communication tool in Russia but international practice proved its ability to influence and engage visitors. For instance, Tate Modern in London, MOMA in New York which have the highest performance on Twitter use it to share link and resources, publish announcements and museum staff commentary (Zafiropoulos, Vrana, & Antoniadis, 2015).

Both Twitter and Facebook give museums a dedicated audience of 100 thousands till 2 mln (for example MoMA page) that are all linked with their personal profiles. Therefore, social networks are amazing sources of data about the museums' most loyal audience that in case of its collection and analysis can benefit museum's marketing campaigns and make interactions with visitors more targeted.

Telling about museum as city space cannot leave apart the tourism value of museum. Evidently, most tourists consider museums as necessary places to visit, often suppose them to shape a common view of the destination or even a whole country. Through museums' collections tourist get know culture and history of the country, find out its traditions and costumes. However, in order to serve a nation museum firstly should be recognisable widely that in fact means be presented in the Internet. The promotion of museums on the international tourism market requires capacities and actions such as the participation of museums in national and international networks or fairs (OECD/ICOM, 2018).

For example, there are world tourist resources as that contain detailed information about tourists' attractions and the presence of museum there guarantees it to be known globally. Museum can maintain their awareness and attendance by regularly updating information about its current activities, selling tickets or reacting to visitors' reviews. The relevant information and positive feedback placed there are able to improve museum performance, while negative feedback or absence of opportunity to leave it are likely to stop people from visiting. That is why presence of museum on such websites should be considered seriously.

Among the most visited travel websites are Booking.com, TripAdvisor, Yahoo.Travel, Expedia are presented (according to eBizMBA Rank 2019). Attractions mostly are searched on TripAdvisor and Expedia. Switching to Expedia and typing for example Paris as a final destination a user open the page with wide range of cultural activities local museums propose starting from "Louvre Gourmet Museum Experience" to "Guided visit for families and children in Museum d'Orsay". At the same time in Moscow there are only three option of cultural leisure that all are related to cities private tours but not museums visiting. As for TripAdvisor, reviews are published there as about experience of foreign museums attendance as about Russian, though the prove of analysis of reviews by museum was found only in interviews of foreign museums specialists. However, the context analysis of reviews would enable museums to build strategy by a such way to best meet visitors expects and requirements.

To sum up, all technologies described in this chapter exist to improve museums main roles fulfilment by accelerating or creating new activities. But at the same time their impact also can be enhanced due to the processing of data they all generate. The opportunities data analysis opens to museum are significant enough to meet the most of museum issues. The third chapter describes their particular value based on the world cases.

Chapter 2. Data employment in museum. Compare international and Russian experience

International experience of data application for collection management

Applying data-driven approach to run museum is a bold but not new idea. Since the end of 20th century the positive impact of data analysis on museum activities has being discussed but only during last six years considerable steps were made to incorporate it. The increasing number of museums around the world has started to understand what significant value data insights would add to their work. According to "Dexibit Limited" over the next two years such aspirations such growing diverse, omnichannel engagement, operational efficiency, governance transparency and funding retention will drive the implementation of data analytics and creation of data governance policy in museums.

Museums roles and technologies described in the last chapter show big range of opportunities museums have to start dealing with data. Around the world a range of examples can be picked to demonstrate the potential of automatic data collection and sophisticated analysis that let museum keeping up the changes surrounding him and better fulfilling its manifold roles.

Keeping the structure of the last chapter we start with example of application of data analysis for conservation of cultural heritage.

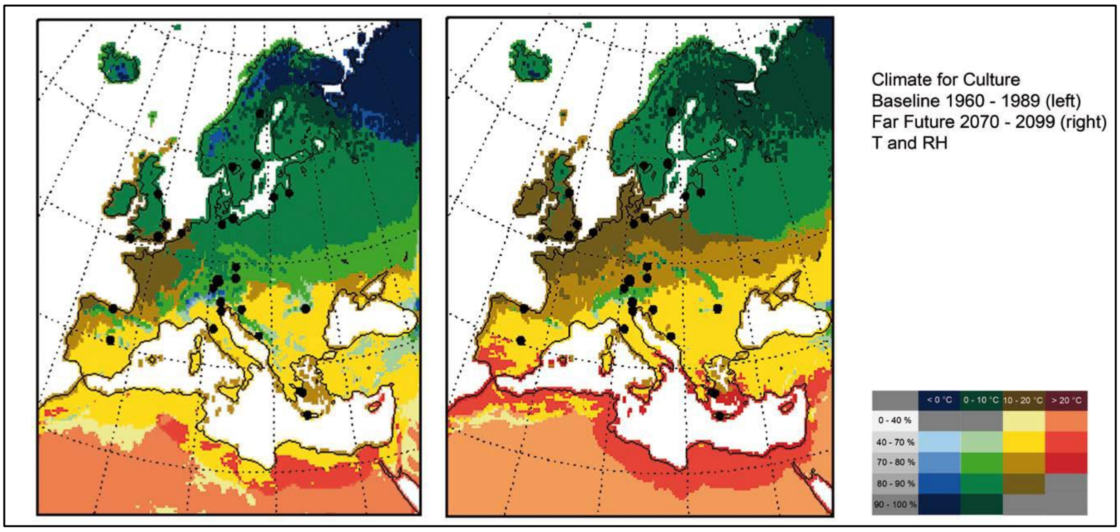
During the period 2009-2014 group of European scientists was carrying out the project *"Climate for culture"* that aimed to explain the effect of climate changes on the safety of cultural heritage (Leissner, 2016). Extreme weather conditions threaten ancient buildings and in-house exhibits to be damaged and museum staff struggle to mitigate the negative effect by maintaining right indoor climate. The goal of this project was to provide conservators with climate change prediction till 2100 in Europe and Egypt as well as point out potentially dangerous conditions that as a result lets them take measures in advance and better meet all uncertainties.

Researches developed a simulation software that contain high resolution climate scenarios on a regional scale changing in dependence of assumptions. For example, scenario A1B that is moderate one is derived from following assumptions:

- rapid economic growth
- increasing global population till 2050, decline after 2050
- rapid introduction of new and more efficient technologies
- balance energy sources

As historic data they took climate records per month during the period from 1960 till 1989. They include information about precipitation, relative humidity, wind and temperature.

Based on this data and considering other recognised climate prediction they create a map depicting all climate changes across the Europe (Pic.3). For example, the model predicts a precipitation increase in the northern Europe and decrease in the south.



Pic. 3 Screen of climate prediction from “Climate for culture system”

As the next step the project translates climate predictions to the state of European cultural heritage, scaling from all region through building and room layers till the sensor layer. In order to solve this tasks, firstly the data from thermohygrographs installed in 106 museums and historical buildings was gathered and combined with regional climate data. Then different simulation models were built in order to find dependence between outdoor and indoor climate changes. Simulations consider hygrothermal transport mechanisms in historic building materials and HVAC climatisation components.

For adapting existing tools for hygrothermal building simulation the building model WUFI® PLUS was used as it is a combination of a thermal building simulation with the hygrothermal envelope calculation model WUFI®. This holistic model takes into account the main hygrothermal effects deriving from moisture sources inside a room, the moisture input of the envelope due to capillary action and diffusion as well as vapour ad- and desorption as a response to exterior and interior climate conditions. In addition, different indoor heat sources coming from the envelope and the solar energy input through walls and windows as well as hygrothermal sources due to natural or mechanical ventilation are incorporated.

As a result of this project, 55 650 climate and risk maps of future induced risks to historic buildings and their interiors were created for each region and type of building. This result is able to better equip museum staff responsible for preservation with knowledge of potential risks and make them prepared in advance.

As the project of particular museum the experience of the National Museum of Ethnology in Japan can be taken into account (Sonoda, 2012). They sought to decrease the number of pests in museum, quick discover abnormalities as well as reduce the consumption of electricity due to data analysis.

In order to identify hotbeds of insect appearance they set traps and daily write down the amount of pests divided by types. Visualisation on the museum map of this points with naming most frequent insects allow staff to clean and monitor these places more carefully. Also as museum collects data for several years they compare results by season and can easily define any abnormalities and quickly take measure to fix problem.

The similar algorithms they apply for monitoring humidity and temperature. These indicators are regularly monitored and added in general graph where all history of indoor climate conditions is represented. That is why all abnormal situations are solved right after they were found thus preventing its negative effect on collection state.

Separately they analysed the change of temperature in several storage rooms with and without air-conditioning and discovered seasonal patterns that showed them when natural climate is good enough and systems could be switched off or turned to the power-saving regime. It did not only save them money but also let them survive in case of electricity shortage for example caused by an earthquake.

In terms of data approach to managing collections there are various cases that illustrate the positive effect, first of data gathering and second of its analysis. Moreover, it benefits not only museum but also government and research community.

As it was described in the first chapter a collection management system stores a vast amount data about artefacts and all what happened to them. All this data is gathered in relation database where in the most cases a provider usually has a direct access while museum uses pre-processed statistic instruments available through user interface. It should be evident such functionality is quite poor to comprehensively manage the collection. That is why some museums choose to develop own collection data warehouses where they aggregate data from CMS database and connect it with other related data. Among the solutions both local storages and cloud warehouses are proposed, however the second way is more beneficial as it makes publishing collection online much easier.

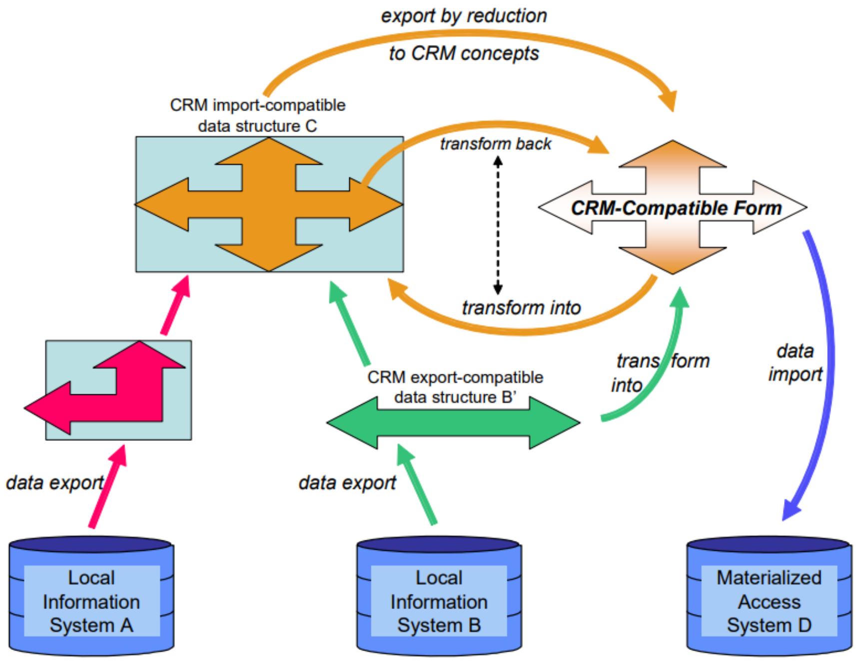
The use of possessed data helps museum run more effectively, however it does not make it a valuable contributor into the development of historic knowledge. The sharing data with the all industry is an essential responsibility of each cultural organisation in terms of its preservation function. Museum around the world came to the concept of “Open data” which means data is placed on open source for reuse, copy, republishing by everyone without restrictions from copyright, patents or other mechanisms of control (Open Knowledge International , 2007).

Publishing data takes several steps if an organisation wants it to be suitable for using. The first one is adjustment of data structure in internal database to standard requirements.

Sharing the data supposes all museums follow the same rules of storing information about their exhibits such as strict set of fields of metadata, image format etc, otherwise it cannot be possible to establish links between internal and external data.

The most popular international standard of recording exhibits data is CIDOC conceptual Reference model (ISO 211727:2014) that provides reference model and information standard that cultural institutions should follow in order to facilitate exchange of cultural heritage information (ICOM, 2011). It aims to make museum information systems *import and export compatible* that includes their capacity to share own data and incorporate external data into museum system in adoptable form. What exactly this form should be is derived from conceptual reference model (CRM). It is an object-oriented model that defines classes, properties and inheritance rules as well as relations between them. Therefore, considering at least the main of them in collection documentation enables interconnections between cultural institutions.

The idea of importable and exportable data structure assumes that the minimum of fields can be uploaded in any compatible system, though other fields can be added or deleted in depend of supposed structure of final system.



Pic. 4 Possible data flow between different kinds of CRM-compatible systems and data structures

The main entities are thing (persistent item), temporal entity, actor, time-span, place, dimension and other. All entities are class instances and have set of main properties. Also some classes have subclasses, for example class Thing has subclasses Physical Object, Symbolic Object, Man-made Thing, Propositional Object and class Temporal Entities has subclasses Events, Period, Production.

CRM is supposed to represent data in structured machine-readable view that first means instances are encoded in formats such as RDF, XML, DAML+OIL, OWL and second all data can be saved in relational database, thus all relations between entities are formally defined in accordance to database architecture rules.
Moreover CIDOC lists main model principals that CRM respects

- Monotonicity – adding extended information to child entity does not compromise relationships with parent entity or property inheritances.
- Minimality – new class is declared only when it could not be a subclass of existent one. One entity can be an instance of two classes (for example Biological Object and Man-made Object)
- Shortcuts rule– properties are shortcuts of longer paths that connects one entity with other via intermediate ones that are not skipped (Physical thing *has owner* Actor though Acquisition(it may not be mentioned)
- Disjointness – classes are disjoint if they do not have connections in real world
- Extensions – model is designed to be linked and augmented with external models
- Coverage – some classes are not thoroughly elaborated as others in order to give a flexibility to institutions to extend the model with other compatible hierarchies to better meet their particular needs. For example, Rights and Actor belong to such entities.

CIDOC community whose members incorporated CRM into their collection documentation practice is very large and includes such museums as Science Museum of London, Australian museum, Brooklyn museum, British museum, National Museum of Denmark and many others. Some examples of exhibit's data are available on the old version of CIDOC official website and allow new member to better understand how it works (for example Epitaphios created by Benaki Museum in Greece (Data Example of the CIDOC Reference Model. Epitaphios GE34604, 1998)).

When data is well-structured the next step is thinking how to open it up. Releasing API and publishing machine-readable datasets are two major ways of sharing data , though they suit to different purposes. While API mainly benefits programmers and web-designers that use it for developing mobile applications and platforms, open datasets are used by wider audience such as scientists, students, government analysts to undertake researches or build statistic models.

As for API development, Powerhouse museum in Sydney was one of the first who understand that publishing data in machine-readable format is essential for a museum who wants to contribute to cultural heritage promotion and historic knowledge development (Dearnley, 2011).

In 2008 they started using the Flickr API and became a first museum who joined their community Flickr Commons. By this time, they had already had an online open public access collection (OPAC) and used the Flickr API to generate tags for image's records (image+metadata)_ as well as link images with associated other images, comments, tags, e.t.c published by Flickr Community. Transporting data from their database to Flickr and then its generated tag back required them to program code on PHP and do check of added data manually.

They found several advantage of being a part of the Community such as advanced information, tags and geo-spatial data related to their images added by other members which museum later used on the OPAC website.

Later museum joined consequently Open Calais semantic tagging API and OCLC's WorldCat API that allowed them to get information about objects mentioned in descriptive texts of images such as people, companies, cities and countries through relating them with other entities related to them.

They ended up with releasing their own API taking into account experience with two previous APIs. What is the most exiting about their solution is that it is used by museum staff also, thus all troubles happen to API or database are identified quickly. Secondly, apart from collection description and its metadata the API provides data about museum services such as cafeteria, shop, membership program etc.

The most of solutions the museum chose to maintain are open-sourced. For instance, they have a Linux platform for serving the project, Python language for writing API and designing its interface with library GraphQL (API Documentation, 2017) as query language that is Facebook open source project. Finally, they chose PostgreSQL as a data warehouse, Apache to run the scripted pages and nginx to serve the static content.

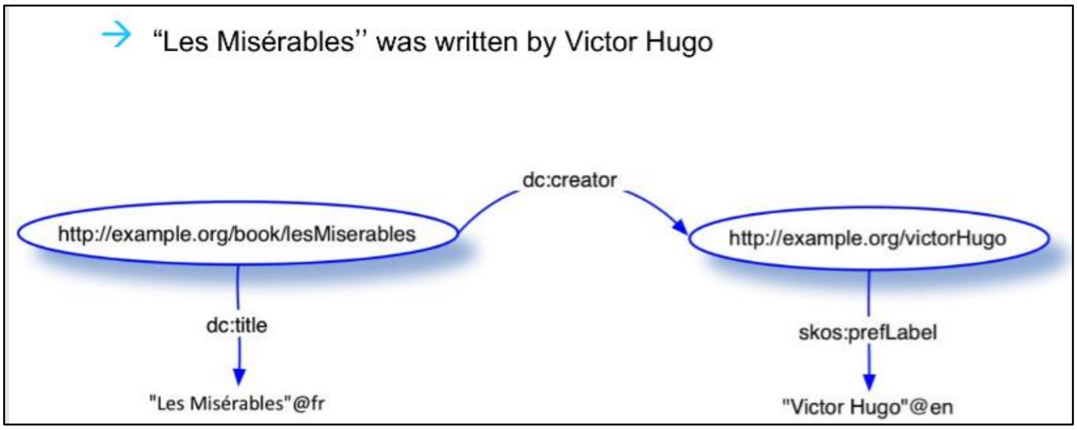
As soon as only application developers use API, museum decided to release WordPress plug-in that allow bloggers add museum images on their websites without writing code.

Finally, museum took attention to API statistic to identify its customer, traffic volume and catch cases of unappropriated usage. Also they constantly augment functionality and fix bugs users face. Therefore, API support is a constant process that requires the museum human resources and significant funding.

Turning to opening data experience, first let's remind the goal of its process. Open data collections exist to facilitate researches and data sharing between institutions. That is why, they are published in format does not demand writing difficult programming code, usually CSV or JSON (Fitzpatrick, 2017). Among wide range of example MoMA, Cooper Hewitt institution and Metropolitan Museum should be named as they were one of the first who made their collections available for public. Their experiences have a lot of similarities such as placing datasets in GitHub, the most famous platforms among programmers and data analysts for sharing open-sourced projects, using Creative Commons Zero as licensing terms that is a base license used organisations to waive copyrights and freely share knowledge, creating data structure in accordance to CIDOC CRM that means it contains at least required minimum of features that the Model prescribes.

Open data by itself does not imply building relations between data but only a free access. Developing a concept further led to creating the concept called Linked Data that is serves to establish networking between cultural organisation and provide free knowledge exchange. As any other concept of connecting data it uses W3C standards that define how website should look and function in order to be the same in different browsers.

Among others they released Resource Description Framework that is standard model that supports data interchange on the Web (W3C, 2014). RDF describes connections between different objects (Pic.5) by noting relations between its semantic representation as well as metadata. It is written in XML. Two nodes (Subject and Object) and link (Predicate) make up a Triple that are in sum compound oriented RDF Graph. Subject and object can be IRI (Internationalized Resource Identifier) or URI (Uniform RI) that is a Unicode string which conforms to some syntax rules, literal that consists of lexical form, datatype IRI and language tag or blank node that are all strictly distinct and distinguishable. Predicate can be only IRI. Any IRI or literal denotes a thing in the world that is called resource that includes physical things, documents, abstract concepts, numbers etc. Each resource described by properties such as author or country.



Pic. 5 RDF triple of CIDOC CRM

The predominant query language for RDF graph is SPARQL that works very similar to SQL as represent the similar structure “SELECT WHERE” but in a bit challenging view. It allows for a query to consist of triple patterns, conjunctions, disjunctions, and optional patterns. According to W3C Recommendations (W3C, 2014), four rules should be followed in order to make data Linked:

— Each object in database should have URI as a key

— Choose HTTP format for URIs

— Provide a guide for a user to help him find URI

— Add links to other URIs in description

Therefore, when data is represented as RDF graph and meet all recommendations it can be easily retrieved by external users without directly querying database.

Yale Center for British Art's Linked Open Data service is a good example of how to correctly provide access to collection data[1]. The browser version allows to test SPARQL functionality without downloading generated files.

With the next query all predicates and objects that belong to the same triplets as the subject with URI "*http://collection.britishart.yale.edu/id/object/499*" are given:

Query: "

*SELECT DISTINCT * WHERE {*

<http://collection.britishart.yale.edu/id/object/499> ?property ?object .

}

SPARQL results:

property	object
rdf:type	crm:E22_Man-Made_Object
rdfs:label	"The Good Samaritan"
crm:PX_display_wrap	"..."
crm:P102_has_title	ycba:object/499/title/1
crm:P1_is_identified_by	ycba:object/499/lidoRecID
crm:P1_is_identified_by	ycba:object/499/TMS
crm:P1_is_identified_by	ycba:object/499/inventory-number
crm:P1_is_identified_by	ycba:object/499/ccd
crm:P2_has_type	ycba:thesauri/300033618
crm:P50_has_current_keeper	ycba:thesauri/department

Pic.6 Part of response of museum collection search engine

In response we see the Subject we ask is the "The Good Samaritan" painting and can follow the link to get further information about its characteristics, exhibitions it participated, bibliography and many others. Also we can see that data is organised in accordance to CRM because properties in the left column have CRM identifiers format "P1_" .

As soon as Linked Open Data supposes interchange of data between institutions we should look again at Europeana project that as was mentioned in the first chapter is a platform aggregates open data of European libraries, museums, galleries and archives. It collects datasets from numerous providers and stores them in their data warehouse developed on MongoDB. Each dataset should contain particular properties, data types and the obligation level of each property Then it converts data to Europeana Data Model(EDM) that is Linked Open Data in fact. Nowadays EDM counts for 36 million cultural heritage objects uploaded by data providers and transformed into linked format. Developing EDM Europeana set following requirements:

— distinguish between a 'provided item' (painting, book) from its digital representations

— Distinguish between an item and its metadata record

[1] http://collection.britishart.yale.edu/sparql/

— Allow the ingestion of multiple records for the same object, which may contain contradictory statements about it
— Support for objects that are composed of other objects
— Support for contextual resources, including concepts from controlled vocabularies

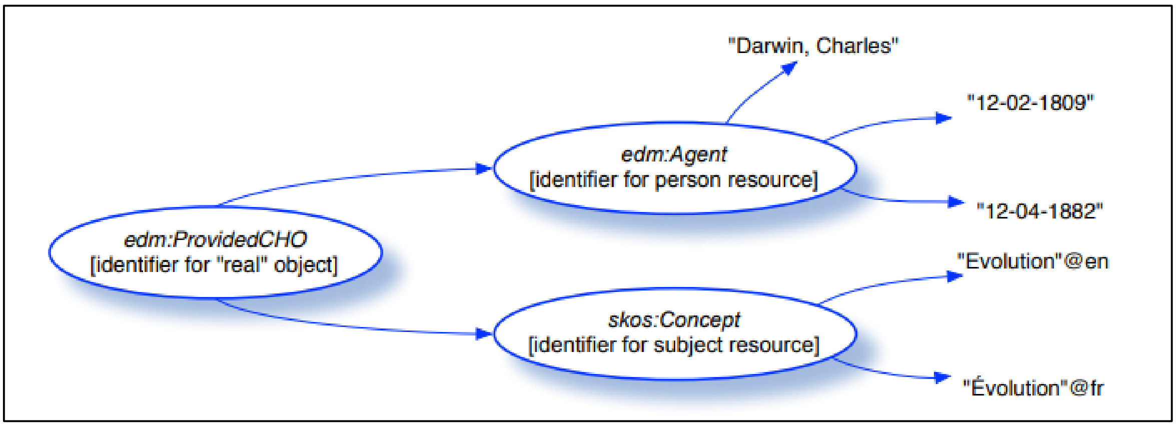
EDM aims to provide more semantics to the data and extend the initial information as possible through specific relations such as "Aboutness", "Part-whole relation", "Similarity" etc.

Currently eight classes are implemented, three *core* of them represent physical objects and five are *contextual* ones.

The core classes are:
— edm:ProvidedCHO - the provided cultural heritage object
— edm:WebResource - the web resource that is the digital representation
— ore:Aggregation - the aggregation that groups the classes together

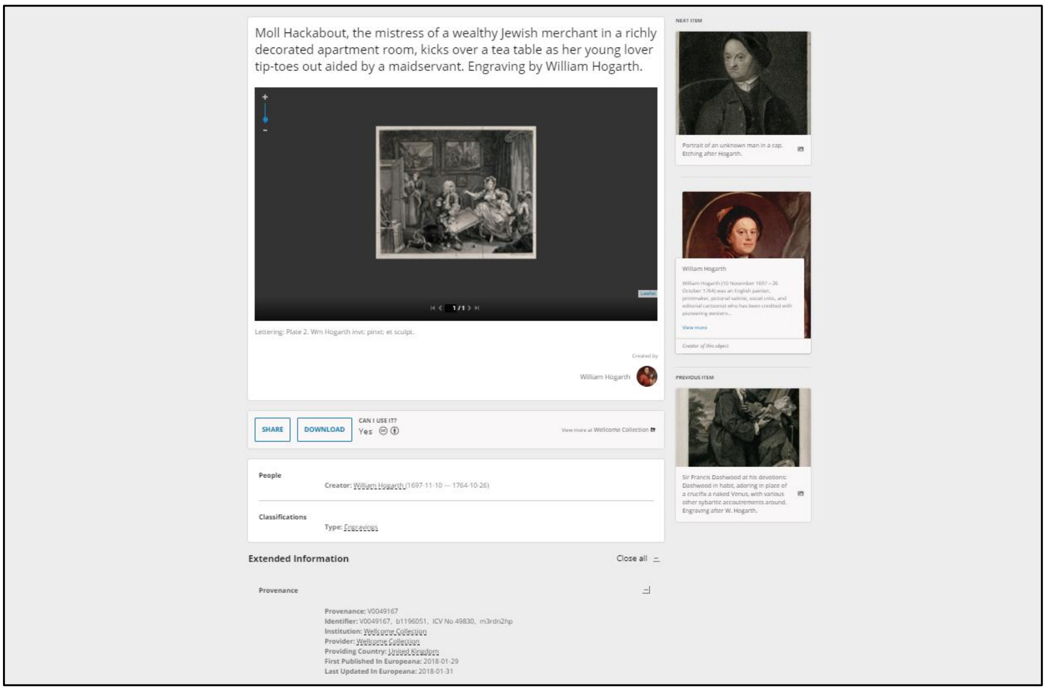
Main contextual classes include:
— edm:Agent – who
— edm:Place – where
— edm:TimeSpan - when
— skos:Concept – what
— cc:License – access and usage

Moreover, EDM is constructed by the way to be able to enrich information about objects from external sources in case if they also publish data in RDF format. For example, in Europeana data model there is a record that the Darwin theory was created by Charles Darwin whose concept is evolution (Pic.7). However, researcher can get more information due to links to the external resources first one about Charles Darwin biography and second about Evolution term available on English and French.



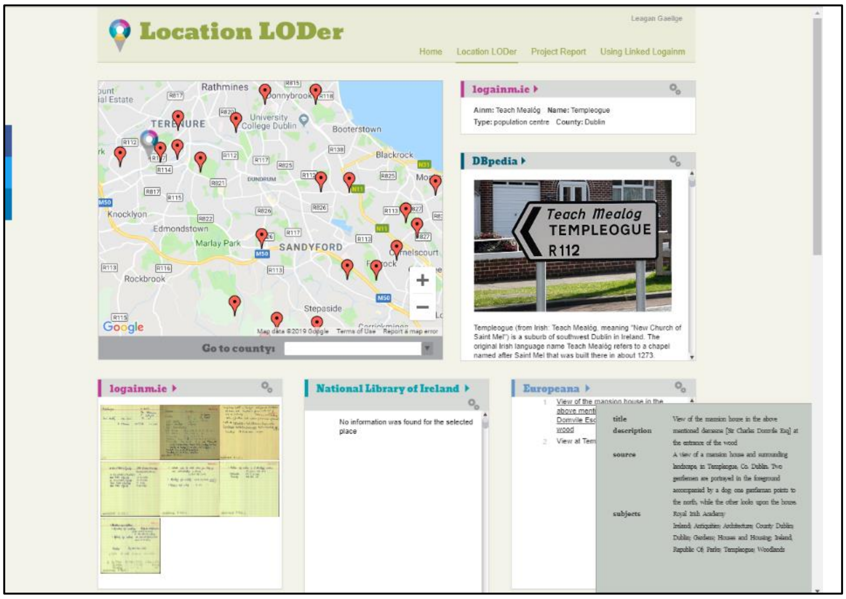
Pic. 7 A Provided CHO with two contextual resources

In order to better understand the advantages of such connectivity the painter William Hogarth was chosen as random object of research. System works as search engine and find all objects related to this name. Moreover, additional filters such as “Media: image”, “Colour: grey”, “Collection: Art” was added. Changing to the suitable page the main information about image (description, technique, date of creation, physical description), opportunity to download, as well as references of the page about its “creator” are provided. In the part of extended information there are links to the pages of data provider “Wellcome Collection”, initial dataset and similar items suggested by Europeana.



Pic. 8 Page of the painting “Moll Hackabout”

Certainly linked data benefits not only search in Europeana website but also projects that use information from the project website in order to enrich their own data. For example, Irish place name database called Linked Logainm that stores information about Irish cultural institution in English and Irish recently has added links to Europeana data on their website that as a result enhanced the catalogues of institutions participate in this projects. All the information is placed on Location LOD website (Pic.9) where each user can get information about cultural object on the map due to established links between geographic dataset and object descriptions from 5 sources.



Pic. 9 Results of picking a place on "LocationLODer"

For example, picking random point on Ireland map the related information from Europeana, Dpedia and Longainm.ie was uploaded that allow to see photo of the place, its history, mentions of this place in literature and some handwritten notes about this place in Irish.

Therefore, the advantages of making data connected to external one are evident, that is why, the increasing number of institutions not only open their data but also convert it to the format available for easy sharing and is struggling to extend housed information by such way.

As soon as museum gathers and structures its data in a collection database a wide range of state-of-art technologies and data analytics methods that bring value to scientific world are widely applied.

Behind the Europeana offers of "similar objects" and suitable handwritten text algorithms of deep learning are laying such as computer vision algorithms that recognise images and help attribute an exhibit to artist or epoch, sentiment analysis that identify emotional state of images and text recognition that help to extract text from handwritten documents.

The most famous computer vision project are belong to Microsoft, Google, IBM and Clarify. They offer powerful pre-trained machine learning models through REST and RPC APIs as well as allow to create custom ones. Among its capacities are images classifying into millions of predefined categories, detecting details of images and reading printed and handwritten text. Finally, it leads to creating valuable metadata into museum image catalogue.

Provider	Create own Model	Tags	face Dectection	Colour Extraction	OCR / Text recognition	Captions	Free Limit	Price
Clarifai	Yes	Yes	yes	Yes	No	No	5,000 transactions free per month	$1.20USD per 1,000 operations. Custom models cost more.
Microsoft	Yes	Yes	Yes	Yes (but only top three colours)	Yes	Yes	5,000 transactions free per month	$1USD per 1,000 transactions. OCR and Face detection cost more.
IBM	Yes	Yes	Yes	Included as part of the tags	In private beta	No	1,000 transactions free per month	Between $0.002 / image and $0.10 depending on classifier used.
Google	No	Yes	Yes	Yes	Yes	Provides a one word 'best guess label'	1,000 transactions free per month	$1.5USD per 1000 transaction.

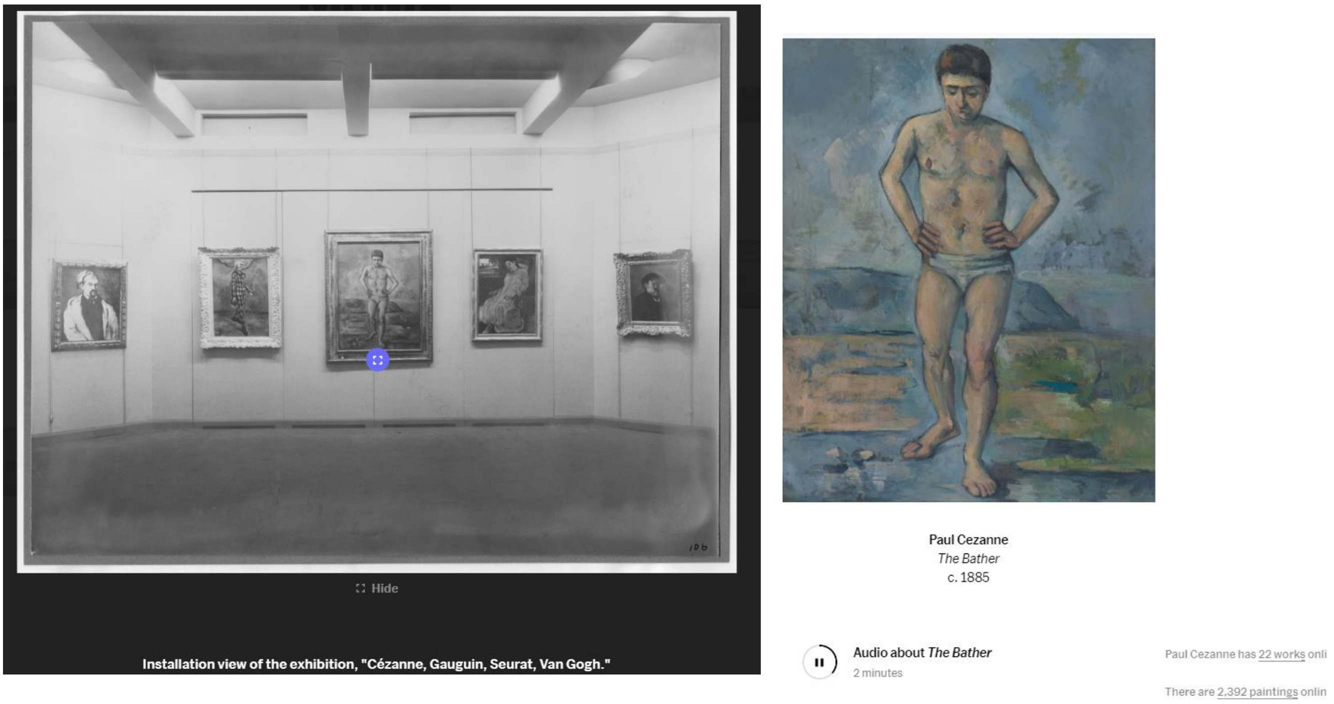
Pic.10 Comparison of computer vision providers

Certainly sometimes some misclassifications happen but in general museum see a lot of value can be extracted from image recognition. For example, Auckland Museum (New Zealand) now is using Microsoft solution to catalogue their collection that is around 7 mln objects (Moriarty, 2018). The system scans an image, identifies objects drawn and outputs Tags, Caption and level of Confidence. When confidence is close to 100% caption is a pretty accurate, otherwise it is totally wrong. The museum chose 60% of confidence as a threshold to automatically accept or reject proposed label. Museum staff use Python script to run algorithms and add Microsoft description and tags automatically to their source system. They point out a significant amount of time that collectors save due to this algorithm.

Image Link	Microsoft Title	tag1	tag2	tag3	tag4	tag5	tag6	tag7	tag8
https://api.auckla	a vintage photo of a group of people sitting in a field	outdoor	grass	person	photo	sitting	group	woman	old
http://api.aucklai	a vintage photo of a truck	building	outdoor	truck	photo	old	man	black	standing
http://api.aucklai	a black and white photo of a truck	outdoor	truck	building	photo	old	road	black	white
http://api.aucklai	a close up of a train	indoor	building	train	dark	black	sitting	looking	view

Pic. 11 Screen of Auckland Museum source system

Another example is a collaboration of MoMA and Google Arts& Cultural Lab (Google Arts&Culture, 2018). The idea of the project was to identify paintings on their old photos of museum exhibitions of 20th century. The algorithm run through 30000 exhibitions photos trying to match drawings in photos with works in MoMa online collection. In total, it recognized over 20000 artworks (over one-third) in these images, so museum used the results to create links between exhibition photos and current collection. The museum added functionality on the website that let visitors know more about paintings exhibited in a particular event by clicking on a sign on the photo.



Pic.12 Matching “The Bather of Paul Cezanne” with its photo on the exhibition of 1929

However this approach has weeknesses such as unability to match paintings in photo with exhibits that are still not added to MoMA online collection or are housed now in another museum. While it is MoMA internal project description will never be complete. For example, in the Pictire 14 from the left side of “The Bather” evidently “The Harlequine” of Cezanne is held but the system did not recognise it because nowadays it is stored in the National Gallery of Art, Washington DC, though they present this painting data in format suitable for linking.

All examples of technologies described above fairly show the high value of data employment in museum. It makes cultural institutions more open to communications and innovations.

International experience of data application to improve visitors experience

As we discussed in previous chapter interaction with visitor, including its education and entertaining is one of museum’s major responsibilities. “In order for visitors to grow and learn from their museum experiences, we need to understand these experiences sufficiently so that we can shape them” (E.Hein, 1998) . Digital services such as audio guides, free WI-FI and VR/AR technologies generate big volume of data about museum visitors that creates the opportunity for an institution to monitor its audience behaviour and facilitates deeper understanding of its interests and motives.

Looking at museums, who care about the visitors’ experience with their exhibitions and additional services, we can see they all gather data about public behaviour that is called Indoor Location tracking. During the visitor-flow tracking they try to find answers to following questions:

- What are the main visitors’ routes in museums
- How much time does it demand to pass them
- What do the key points exist along all tracks
- Are answers to previous questions similar to curators expectations

There are a wide range of ways to get data about visitors’ movement and each museum chooses one that suits to its level of technical equipment and meet specific requirements.

The first way of collecting data about visitors that comes to mind is taking data from audio guides. As visitors carry with them during their whole visit and switch on only when see something interesting the data from these devices is able to tell a lot about visitors’ experience. Turning to Met experience with audio guides, first thing they found out due to these devices is who is its audience that implies which countries they from and what is a purpose of their visit (Grace, 2015). First insight derived from the language a visitor chooses in device, another is from a set of options he chooses there. Also audio guide helped museum staff understand challenges visitor meet during exploring collection that led to first simplifying the numbering system of exhibits, signage, logos and other supplementary materials and second to making changes in audio guide itself, for example rearranging some of offered tours.

One of the most popular means is IBeacon technology developed by Apple. Beacons are small wireless sensors that communicate with Bluetooth-enabled smart devices such as iPhones or iPads that constantly declare their location using a Bluetooth low energy radio transmitter. In turn, smart devices monitor the received signal strength indication (RSSI) and determine the device's proximity to the beacon. Once the user is in the desired proximity range, the corresponding app content is triggered.

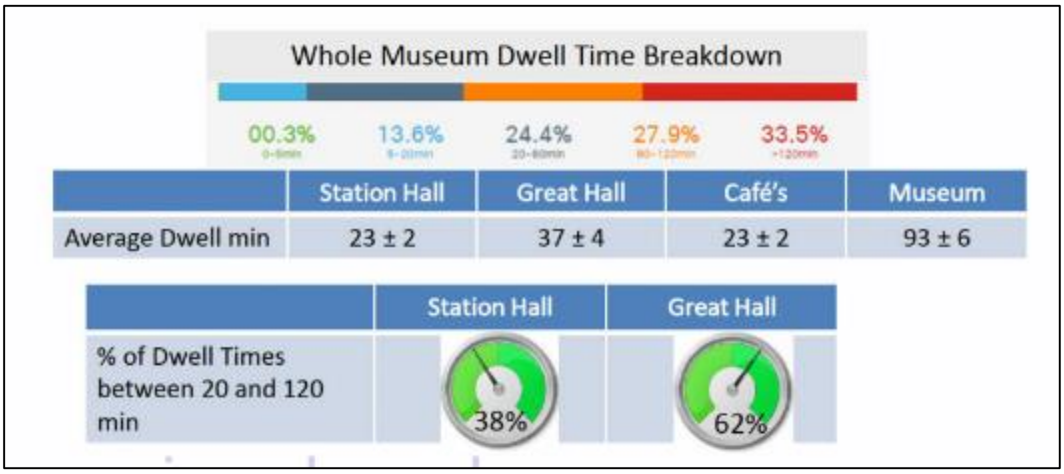
As for museums, Beacons offer them an opportunity to provide context to visitors. It is considered modern alternative to audio tour because dealing with IBeacon guest does not have to search relevant track as it defines it by itself based on visitor’s location. This approach let visitor to learn more and stay more satisfied about his attendance. App can also send recommendations based on the exhibits and artefacts that a visitor has expressed interest in during their previous visits.

The technology has already been implemented in wide range of American museums such as The Metropolitan Museum of Art or Brooklyn Museum as well as in European ones - The Neon Muzeum, Poland, Rubens Art Museum, Antwerp or Groninger Museum. For example, The Art Institute of Chicago has leveraged its 300-beacon network (activated when visitors connect to WiFi) to increase paid attendance from $14.8 million in 2015 to a projected $19.9 million the following fiscal year. The user data is crunched to produce coveted analytics like “heat maps” (a visual representation of the total number of interactions with an exhibit), travel paths (where people walk) and dwell times (how long they stand in a gallery). Such insights help curators better understand what places and exhibits attract visitors the most and what places are skipped. Possessing this knowledge give them opportunity to redesign exhibitions or release more descriptive material in order to make educational process more pleasurable.

But it has some disadvantages and restrictions. First of all, Jennie Browne from Brooklyn Museum notices (Browne, 2014) that beacons make mistakes when detect the location, because of the weakness of signal. Bluetooth technology is based on radio waves. Radio waves are inherently flaky because of interference that in this case can mean any work of art, vitrine, installation partitions, or even other visitors in the gallery. Ultimately each gallery has its fair share of radio wave blockers. So it makes beacons not very reliable sources of visitor's data.

The alternative source of visitor movement data in museums is WIFI network. Now almost all big museums offer the opportunity to connect to their WIFI and nearly 40 % per cent of visitors do it or have their WIFI turned on. In both cases information is gathered.

Connecting to WIFI creates two data flows. First is their personal information as phone number or email that people enter in order to get access to network and users cookies he creates surfing on the internet. This type of data not only allows to understand who is our visitor in each exact moment but also afford analysts realize that a visitor come back to museum. Second flow is exactly visitor's movement on exhibition, that is received from WI-FI trackers installed in each second hall. Taking again the Art Institute of Chicago experience we can see that data from the museum's Wi-Fi network really can help answer business questions. In their case they find out that visitors follow a similar path through the museum regardless of where they enter. They have also learned that using exhibition-like interpretive techniques in the galleries displaying our permanent collection significantly increases dwell time there.



Pic 13 Example of dwell time statistic gotten from WIFI system

Apart Art institute of Chicago visitor tracking analysis exists in at least three of Britain's most popular cultural institutions the National Gallery, Natural History Museums in

London and National Railway Museum in York. They have all revealed that they had tested or deployed tracking software (Malley, 2017) which could conceivably help curators and managers make decisions. For example, the Railway Museum's analytics initiative revealed that most people spend around 93 minutes hanging around the museum, while there are 33.5% of 'uber-nerds' visitors who spend over two hours looking at trains. Museums analysts also confirmed that data could also show the routes people took around the museum - with this slide showing where visitors to each main hall came from - and where they went to after.

The older and more comprehensive method of tracking of visitor's routes is analysis of data from video cameras. Technological advancements in video equipment, such as digital video and video editing software give visitor researchers a valuable method to document visitor behaviours for study. Benefits of collecting visitor behaviours on video for study include: advantages when working with dynamic visitor behavioural data, especially during coding; video saves time and video decreases visitor reactivity to researchers and data collection tools.

Unlike WIFI and IBeacon tracking video analytics capture a visitor's movements and mimic that define his emotional stat and allow to realize his reaction to what he is seeing, feeling or listening. Moreover, the accuracy of visitor location appointed by cameras is higher than by Bluetooth and WI-FI trackers. On the other hands, it fails with preserving visitors' identities because it requires system to make face-recognition of hundreds of people in real-time that is almost impossible considering network capacity. This method is quite expensive in compare with two others and now is mostly applied to provide safety of visitors and collection rather than for analytics purposes.

A lot of information about visitor's reaction on what they see around them can be given from AR data or special eye-tracking glasses. Recording the gaze brings insight into how people look at objects. Based on these findings curators can change exhibits position. For example, during the visit of Glasgow Kelvingrove Art Gallery and Museum one sign in the hall of impressionist took attention that claimed paintings were lowered a bit because such height guarantees a better perception of images. Such conclusion was likely to be made due to eye-tracking.

The experience of Van Gogh Museum in eye-tracking says that the museum not only corrects objects' positions but also tunes educational programs and tours to show such details that usually do not draw visitors attention.

International experience of data application for communications enhancement

Being a public place involves museum into communication with people. Since people belong to different social groups and expect different outcomes of museum engagement, there are numerous purpose of possible communications. Hence, interactions with tourists aim to introduce them culture and history of them place, while relations with locals imply social activities and educational programs. Moreover, first relations are inherently temporal and can hardly be prolonged, whereas museum should seek to last second as long as possible.

To keep audience in the long run, museum distinguishes communications with people who have already visited museum from people who have not had such a chance yet. More smart clustering of visitors' flow is an output of data analysis. For example, in Denmark cultural industry divides the audience into several groups such as 'recharger', 'facilitator', 'explorer' that derived from exploring visitors 'motivation and learning behaviour' (Jensen, 2015). Assigning each person to one or several of these groups helps museum to individually approach him and create personal promotion and communication strategy.

Modern technologies, as it was described in the second chapter, meet museum's needs to establish contacts with every person. Marketing instruments allow museums to create personified strategies and influence visitors when they are physically far away from museum building. However, the highest efficiency cannot be achieved without data analytics at the core of all initiatives.

First of all, the museum's online presence, including website and social networks pages mean big engaged audience, that deliberately shares their data with museum. The analysis of their "data trails" through museum resources give an opportunity to develop personalized strategy of interaction which potentially can result in building long term relationships.

The best examples of social network analytics were shown by MoMA (3,6 mln followers in Instagram, 5,57 mln in Twitter), MetMuseum (2,4 mln followers in Inst, 4,44 mln in Twitter) and Louvre (1,6 mln in Instagram, 1,4mln in Twitter). Their work with social metrics of public includes: Facebook analytics, Twitter analytics and listening report, Instagram analytics and management, LinkedIn company pages report, social media analytics dashboard, review and opinion sites, bookmarking, media sharing, and more. Social data analysis in Met Museum provides good example how applying social media metrics promotes better understanding affinities of people as well as what people want when they interact with art. They examine Facebook and Twitters accounts of their followers to determine individuals who are more likely to respond to their marketing campaign or will be interesting to join some event. Also they analyse selfie behaviour in terms of personality and preferences in art. Also they pay attention to the frequency of visits per year. Additional advantage they emphasize is that data allows them to cluster clients and personalize interaction with them.

Another interesting example of social media analysis in museum was provided by Tate Museum (UK) during the festival "The Tanks: Art in Action" hold in their new space. The program presented in these industrial spaces featured an array of installations, live performances, and films. During the all festival time they communicated with audience and promoted events via Twitter. It would have seemed ordinary case if they had not created "Comments Wall" right in exhibition hall where in real-time each visitor could see his comment about the festival that leave in Twitter with hashtag #thetanks. This idea encouraged visitors to leave tweets and made it possible for museum to collect huge amount of textual data that allowed them to understand people's reactions and thoughts about new exhibition space. To get these insights, first of all museum extracted information from metadata as age, gender, location and number of followers to identify person profile and realize if he really came to the space. The main part of analysis was dedicated to the context and sentiment analysis. The content of tweets was analysed to recognize the authors aim, tweets topic and sentiment whether it is positive, negative or neutral. Also with data they measured the effectiveness of the marketing and communication campaign examining volumes of tweets per day, the potential reach of the news, and the traffic generated to the Tate's website.

The increasing attention is given to work with youth audience that is more demanding of interactivity than older generations. Actually social networks, video/photo sharing sites, online games, devices like smartphones and Apps are now mainstays of youth culture. In this scope, museums find useful to scrutinize why young people find these activities so interesting and important and apply this for engaging them. According to research "Living and Learning with New Media: from the Digital Youth Project", and the paper "Digital media and youth: social uses of new technologies" teenagers have lots of advantages from involvement in social media life. They are able:

- to increase friendships, interests and independence
- to extend or strengthen social relationships
- to acquire social and communication competences
- to affirm the own identity, by communicating ("post", "like" and "share") or, for instance, customizing their own social page
- to be engaged in peer-based and self-directed learning: this allows to acquire various digital skills, make and share own creations and receive feedbacks from other young people online

While examples above target people without regard their location, the next analytics cases mostly touch citizens who is able to frequent museums. In the previous chapter, it was mentioned that last years in order to benefit the most loyal visitors museums have been creating special programs called Museum's Friends to which every keen person can join. Since these programs demand each member to leave personal information and assign to him a unique identification number it becomes easy to museums monitor members behaviour and gather data about their activities from ticket sales systems, CRM system and the websites where museum public oncoming events.

Carnegie Museums of Pittsburgh (CMP) provides the example how to get data about members of the loyalty program and uses this information to increase effectiveness of marketing activities (Lin, 2016). They ask visitors about their interests when signing up for memberships as well as draw insights from member's purchase history. According to the Senior Director of Donor Relations and Membership, "her team works to make sure that the messaging to all members makes sense, and that members are not getting too many emails in any given week". The goal is to use these indicators on the customers' records to create a flow of information that's both logical and manageable. The effectiveness of such collaborations is appeared through members responses that imply both comments and feedbacks in social networks as active participation in museum social and educational activities.

While loyalty programs build long-term relationships with locals, museum also should think how to better fulfil the role of tourist venue. Their high-quality online promotion leads to not only their positive image but a whole destination. Social media and tourist online services that enable tasks performance and data analysis can improve its quality (Jensen, 2015). Danish Agency for Culture claims that demographic indicators taken from Facebook pages of Danish museums followers clarify the statistic of tourists' countries of origin and as a result adjust supplementary materials and even exhibitions design to this demographic.

The important part of tourist attraction strategy is a analysis of reviews but now the most of museums process them manually, hence slowly and not systematically. While there are lot of scientific papers and articles (Zanibellato, Rosin, & Casarin, (2017);Alexander, Blank, & Hale(2018); (Maurer, 2011)) describing machine text analysis, only the British museum in London implemented it into its marketing strategy on a constant base (Cuau, 2018).They launched collection of visitors feedbacks and enquires in 2016 with a goal to "know what visitors think in real-time". Insights identified problems with readability of information about luggage policy, so museum made it more prominent and added it to pre-visit emails that both increased awareness of rules among visitors and help them avoid unpleasant moments during the visit.

As for exploration of TripAdvisor reviews, they first made it manually. But given 1000 reviews about the Museum published every month they decided to query TripAdvisor office and ask them an API access for automatic download. Finally, they were allowed to upload two years of data featuring dates, titles, languages and group type. Due to Natural language processing (NLP) algorithms they extracted key words from texts and grouped them by topics. Also they undertook sentiment analysis that divided all pull on positive and negative reviews. As additional metrics, they used languages, number of people and a ranking score. This method was found a quite effective, that is why they distributed it through other review platforms such as emails, social networks, museum website and local analogues of TripAdvisor. Finally, based on this data they created a cross-platform satisfaction score that they would track over time. The most pleasurable thing of this experience is an open-sourced code that can easily be distributed to any museum that wonders to try.

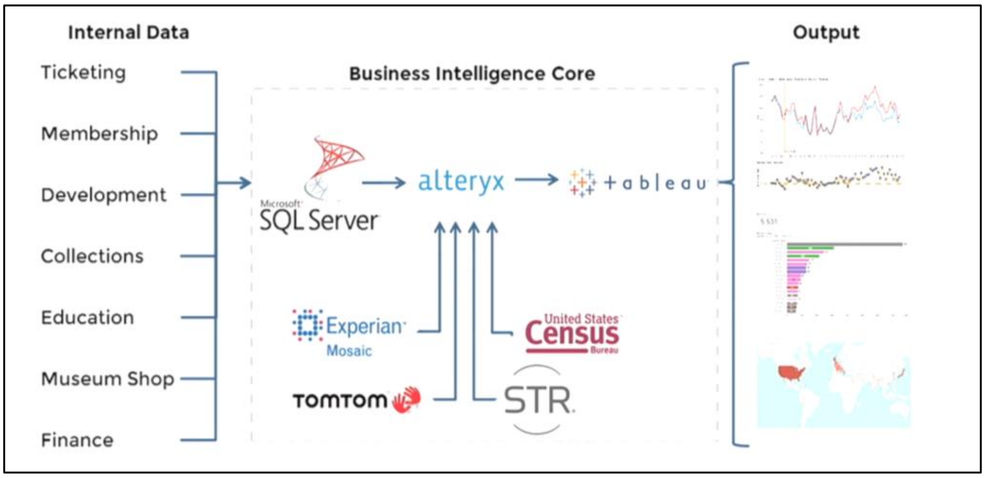
International experience of systematic approach to data collection and analytics in museum

All examples given up to this point describe the separate methods applied in museums to benefit its particular missions. Here, two cases of complex approach to management are provided, though they have some gaps that also will be highlighted.

The concept of developing the unite data infrastructure within one museum now is being implemented by Art Chicago Institute and British museum. The first one created "business intelligence core" that let them apply data to decision-making across the organization (Simnick, 2017). They extract data from ticketing systems, collection management system, database of museum shop and other internal sources.

As data storage they created a central data warehouse managed through Microsoft SQL Server. All data is processed via self-analytics software Alteryx. It does simple machine learning , automates structuring raw data in dashboards, visualize outcomes in Tableau BI-system Tableau. All this big and complex infrastructure is supported by two people.

According to the museum analyst, it has allowed museum to develop insights about their issues. This has the greatest impact on the Art Institute. For museums who is new at applying data analytics, it is also an effective way to build capabilities without making large investments in staff or technology.

- 

Pic 14 Concept of Art Chicago Art Institute's data architecture

However, such approach has evident drawbacks. First of all, the support of such composite system demands additional human resources and their time that a museum usually does not possess. Moreover, it can hardly be scale or augmented because current solutions do not suppose processing big amount of data as well as desire to add a new source requires to rebuild all infrastructure.

What about British museum they are building data infrastructure to address business problems in real time. They want to implement data-driven approach in each process of making decisions. By now they successfully built a minimum viable product of the Big Data pipeline bringing three data flows related to visitors' insights: visitor counting, audio guide visit data and Wi-Fi data. They centralize these flows in data warehouse that enabled museums to facilitate at speed analysis. Main tool British museum applies to process data is Azure Technology, to analyse they use R and PowerBI. Currently they are in stage of getting data from external sources and consolidating it with internal one in one data pipeline. The case of dealing with analytics in British museum can be considered the most mature.

Since the understanding of museum role in social life is increasingly extending, the variety of external sources whose data will be useful for museum work is growing. Much wider opportunities for big data analytics start emerging when cultural community firstly open all data exists for all its members and secondly integrates with other communities and services that are not directly related to museum activity. Unfortunately, small amount of museums now is including external data in their strategy, though public reports often notice its benefits. For example, the report of the American Association of Museums call museums upon using statistics of the demographics of local communities from U.S. Census Bureau. Due to it sees museum role as driver of local minorities engagement in cultural life, it is necessary for museum to know what communities exist in surrounding area and who is their target audience. Also the report highlights the increase of creativity that museum can reveal in activities for youth generations perspective due to the awareness of their social conditions and cultural habits. Such information is stored in databases of state specialized institutions and is openly published. That means there are no limitation for museums to use them except own technical capabilities.

To sum up, during the last fifteen years a significant progress in data processing was made in museum sphere that is proved by international examples provided above. However, the two main gaps were discovered. First of all, the most museums use data in order to solve some particular issues rather than completely transform internal processes into data-driven. Next, when museums need data insights they only access to internal sources, however open data sources as well as data of other institutions related to cultural contain lots of useful data. According to Dexibit specialists, the connections exist, for example in telephony, when client's data can be easily transported from one mobile operator to another represent the future of identity for museums whether visitors experience or collection' objects perspective.

Russian museums realities in data management

Russian experience of data analytics in museums is not so significant and systematic as international. The work with data usually ends on user's statistic reports that are uploaded from ticket or collection systems. Such reports have a restricted number of fields and poorly service museums with significant data insights. The general technical solutions such as TicketNet system for accounting visitors and KAMIS system for collection management do not allow museums to use data for decision-making.

The audit of IT-infrastructure undertook last year as a coursework in the Pushkin Museum of Fine Arts revealed almost full disconnection of internal data stores that causes impossibility to automatically build reports and gain insight on regular basis. Internal databases have problems such as inconsistency, decentralization and what is more, limited direct access to database for museum staff. Certainly, without a holistic approach to data processing no real-time methods can be applied.

The weak attention to data causes challenges in all museum spheres. First of all, Russian museums cannot participate in global cultural knowledge network because it requires accommodation of collection database to CIDOC CRM that is not introduced in Russia. Also, there are some issues with client relationships. Museum friends when they buy the card are promised to be treated individually, for example be notified first about coming events. However, according to reviews in social networks members often do not even receive public notifications not mentioning personal ones. Lastly, museums struggle uneven visitor flow distribution that results in empty halls in week days and overcrowdings in weekends.

Despite the weak current data infrastructure in Russian museums, the cultural conference "Digital Museum. Innovations for museum and exhibition spaces, 2017" confirmed the growing interest to data analysis in cultural industry. There some museums presented test projects of applying analytics techniques into daily routine. For example , the CIO of the Pushkin museum of Fine Arts Vladimir Opredelenov described a video-analytics project museum had carried out before. They used computer vision algorithms for people recognition in order to create visiting heatmaps (The Pushkin museum of Fine Arts & JSKT data group, 2017). Such heatmaps are used globally to analyse visitor distribution across halls and better organise paths for tours, booklets and audio guides. However, the project was not implemented in museum's regular practice because it required manual data collection.

Existing restrictions related to museum state status and resource limitations do not allow museum to apply any of the state-of-the-art technologies that are spread across private companies. For example, museums cannot store personal data on Google clouds because they are located outside Russia. It means such analytics tools as BigQuery is not available for Russian museums, though American museums benefit a lot from their simplicity and effectiveness. However, it does not mean that it is impossible to build data infrastructure in Russian museums at all.

The importance of automatic decision making in museum was pointed out by the deputy director of Politech museum Michel Yakovenko (Savina M, 2017). In his speech he took attention to the idea of data exchange between museums, government and aggregating platforms that can be brought to life due automatic data collection. The open cultural data of the Russian Ministry of Culture with some open statistic of museums performance and “Goscatalog”, where digital version of collections from all Russian museums is being gathered, are named as big data sources that can boost data approach in Russian museums.

The idea is reasonable because these projects first, require museums collect data and second, make it open for users favour. However, there is neither plan how museum should enable data collection nor standards to regulate data structure and metadata. Therefore, in the best case museums set up their own rules for data presentation or they do not consider any rules and send data manually in inherit view. For example, Goskatalog does not standardize well the format of images and its metadata. That is why while objects of one museum can be explored in details, another ones provide images where only main traits can be recognised. Kamis system partially solves problem with upload to “Goskatalog” because automatically send data stored in systems database to government storage, however not all museums use this system first and secondly, the provided format does not meet requirement of global cultural knowledge sharing.

Lastly, every museum as a state institution accords their target KPI with authorities when they plan budgets for next year. Today plans are set without regard to data insights that make them sometimes not realistic. Moreover, government monitoring is difficult to undertake because information provided by museums contains only general numbers that do not allow to find the cause of museums difficulties.

To sum up, Russian cultural industry faces a lot of challenges in terms of communications and management. Index is an instrument that is able to address them. It will help clarify, rebuild current processes and boost the development of new ones. Moreover, not only museum themselves but also government will benefit from it.

Chapter 3. World Big Data best practices relevant to museums

Best practices of Big Data adoption to problem solving

The museum experience of dealing with data is much less significant and systematic than this of other industries. The emerging demand for deep data analysis supported by increasingly high volume of available data turns enterprises to Big Data technologies. Cloud storage and real-time analytics have become an inherit component of running business that increases operational efficiency and improves decision-making process.

Telecom companies and banks are the leaders among offline organisations. They use data insights to better run their branches, improve customer relationships and monitor the state of cell towers and other equipment. As a result, speed and efficiency of connections between all parts of their business and essential to them. So they make efforts to enable the highest possible speed of information exchange in their network. Their experience proves a positive influence of data-driven approach on the whole business performance.

Developing data approach for museums cannot miss the reflection about practices that can improve museum business process. That is why in this chapter world-leading Big Data solutions that potentially meet museums needs are discussed. Each case has been chosen to show an alternative solution for fulfilling one of three museums roles. All of them were chosen due to the similarities either of their challenges or key goal with museums.

Non-commercial application

As soon as collection data exchange implies graph model application, first look is given to the use of graphs. The platform *"Graph Commons"* is a vivid example of using Big Data for social reasons. Being a collaborative platform for mapping, analysing and publishing data-networks it was countlessly used to highlight issues that cause ecological or social damage in some country or topical political researches about current world events. For example activists in Brazil have mapped public-private partnership causing deforestation in Amazon rainforest and Zuirch-based NGO monitors lobbying influences in the Swiss parliament. Moreover, platform motivate organisations including art institutions to unite in hubs , share data and build networks together.

On Graph Commons members can collectively compile data about topics of interests, define and categorize relations, transform their data into interactive network maps, discover new pattern and share the isights about complex issues using a user-friendly interface. The platform serves both producers and consumers of graphs by linking entities together in useful ways and thereby creating a whole that in greater than the sum of its parts. Also, once the data is imported, network analysis can be applied to discover indirect relations and organic clusters that are otherwise hidden. Moreover, opened links and patterns can be used as a features for future deep analytics.

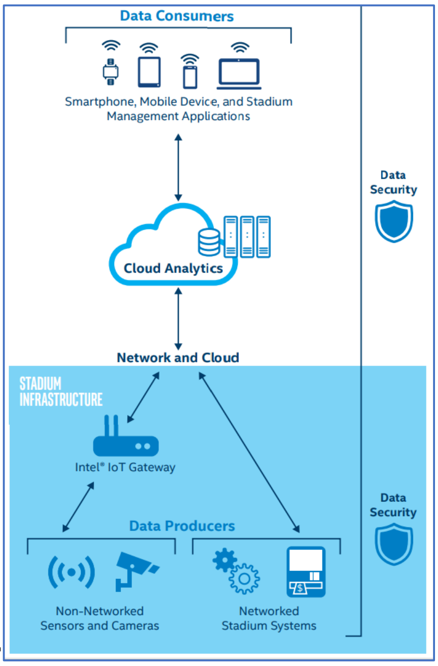
Smart home solutions

Within the environmental control museums do building and collection monitoring that can potentially be automated through smart home technologies. Smart home is based on Internet of Things (IoT) that means it is an innovative solution for managing building with network of sensors that collect environmental and behaviour data, central management system where they are controlled and regulated and wearable devices that get information and present it in user-friendly view. Holistic Smart home encompasses energy system, facility management system and audio-visualize system. Smart home main objectives are energy conservation, security, environmental protection, consumption control and other.

Wireless sensors that include cameras, sound sensors, smoke and light sensors, atmospheric pressure and temperature sensors record events and transmit them to *Intelligent gateway* that sends received signals to the Internet. Then data is stored in data *cloud warehouse* where it is processed to be sent to home appliance and personal devices of home residents. Also there are networked sensors that collect home internal and external data and continually transfer it via the local network, to the smart home server.

A collection of software components wrapped as APIs, allowing external applications execute it, given it follows the pre-defined parameters format. Such an API can process sensors data or manage necessary actions.

In terms of visitors tracking the museum can learn from stadiums. Large museums as overcrowded stadiums should know how to smartly manage visitor flows and prevent queues and shortage of resources. In 2016 Intel introduced a solution for football stadiums that looks suitable for museums as well (Intel Corporation, 2016). They provide a platform that is aiming at improvement of operational efficiency, fan experience and stadium security. The solution's architecture looks similar to Smart home application however the main accent is set on crowd control and fan loyalty increase. Outfitted with numerous cameras and sensors stadiums feed real-time data from restrooms, concession stores, gates and corridors that trigger such actions as washing toilets, food restock, search for lost children, preventing fights, parking surveillance, catching fun faces on stadium to broadcast on the screen. Also they release mobile applications that first augment watch game with eye not catching details and second allow to stay informed about stadium and team events and offers. These applications serve to collect personal fan's data and target advertising. In order to make data boosts actions it is ingested from sources, sent through IoT Gateway to the Cloud, analysed there and then distributed across mobile devices and Stadium management applications (Pic.16).



cr

Pic. 16 Smart Stadium data architecture

Also stadium applications are equipped with analytics tools that allow stadium staff to visualize trends and build long-term strategies.

The social role of museum makes him a part of a city environment that should think how to communicate with other public services and employ their practices in cultural field.

Smart city concept

First of all, let's look at Smart City concept and its advantages. At the core of typical Smart City project there is data analytics storage where all city data is aggregated, thus providing each public service and local authority with structured data available to improve their performance. Moreover, this data is usually openly published online to let each person use and manipulate with this data for free. Open data should be machine-readable and easily accessed without manual activities. Since no legal restrictions exist to access this data, every service can use it to enrich own data. The most recognisable ideas of projects that benefit from city-wide data nowadays relates to transport services. They collect data across all mode of public transport such as metro, trams, buses as well as private cars and bikes in order to manage traffic. For example, the City of Copenhagen now is under development of solution to control flows of cyclists and pedestrians in order to improve accessibility and walkability of routes and public transport stations. They want to understand what obstacles they meet and how to make roads more comfortable for non-vehicle traffic members. The solutions for extracting data are categorised on tele data, WI-FI sensor, Computer Vision Sensor, Transport Survey App. They analysed each source and detected their pluses and restrictions. For example most of them are time and energy consuming, though the collected information is real-time and comprehensive that make subsequent analysis more accurate. Other examples of Smart city solutions help waste collection, firefighters and police to choose optimal itineraries to reduce time and energy costs.

Public places also can improve social environment with from public data, though today all solutions either work with statistic historic data or with users data aggregated on mobile platforms rather than organisational one. However, it does not make projects less significant and perspective. The pilot project within initiative Open4Cities in Rotterdam shows how parks can use data to make citizens leisure time more pleasurable and socially active (Slingerland, 2017). Within the project a digital platform that builds park communities was created. It analyses park users profiles and matches them with other visitors based on their skills and interests. In park there is a wall of activities that are organised as park coordinators as users and each user can join any of them. As a result, matched visitors are likely to attend the same activity and get know each other. Activity is going if there are enough people matched to it.

Moreover, park also benefit from this application as it collects demographic statistic and get realized how are its visitors and what interests they have. For example, park use this data to confirm its contribution to society development in front of the municipality and get enough investments. Therefore, simple enough technological solution assist park with fulfilling its community building function and museums can learn from them

With solutions for locals, museums should not forget about other consumers that usually have only one physical contact with museum however often keep in touch online. Tourist engagement is a very important museum function that can be achieved more sustainably and efficiently when data experience of other tourist market players is considered.

According to the current practice, social media platforms constitute a source of valuable information for players of tourist market (Brandta, Johannes, & Neumann, 2017). First of all, they make a tourist itinerary evident because almost everything person publishes on his page such as tags, comment, location coordinates and even photos allow to identify what he visited and in which order. Semantic analysis together with computer vision extract significant information about places, emotions, plans from his comments, review, posts and photos. Tweet and review sentiment analysis is the most spread method to complete tourist vision about organisation and check the level of satisfaction of particular person There are next steps: extraction and aggregation of inherit text into dashboard, pre-processing to transform data into machine readable format, feature engineering to enhance data with new mathematically calculated features, model building and interpretation. As sentiment analysis algorithms support vector machines and Naïve Bias Classifier are often used that in case of two classes return either positive or negative.

Moreover, Smart City solutions also generate valuable for tourist analysis data. Data from mobile devices is a valuable source of tourist movement information that each territory struggling to know as more as possible about touristic behaviour wants to possess. Information can be aggregated mobile operator (call detail records) that catch the signal of data coming through its networks or translated by device itself. Traffic from scenic spots, call detail records, geolocations in social networks, signalling data of mobile phones matched with GIS data tell about full tourist way what he sees, tastes, enjoys or suffers. The most of this data is conveyed and stored in real-time that as a result demand using Big Data applications such as Apache Hadoop or Spark.

Similar to mobile devices credit cards are able to convey information about tourist and his country of origin. While personal information is not disclosed by banks the coordinates of places or online sources where a tourist uses cards can be reached.

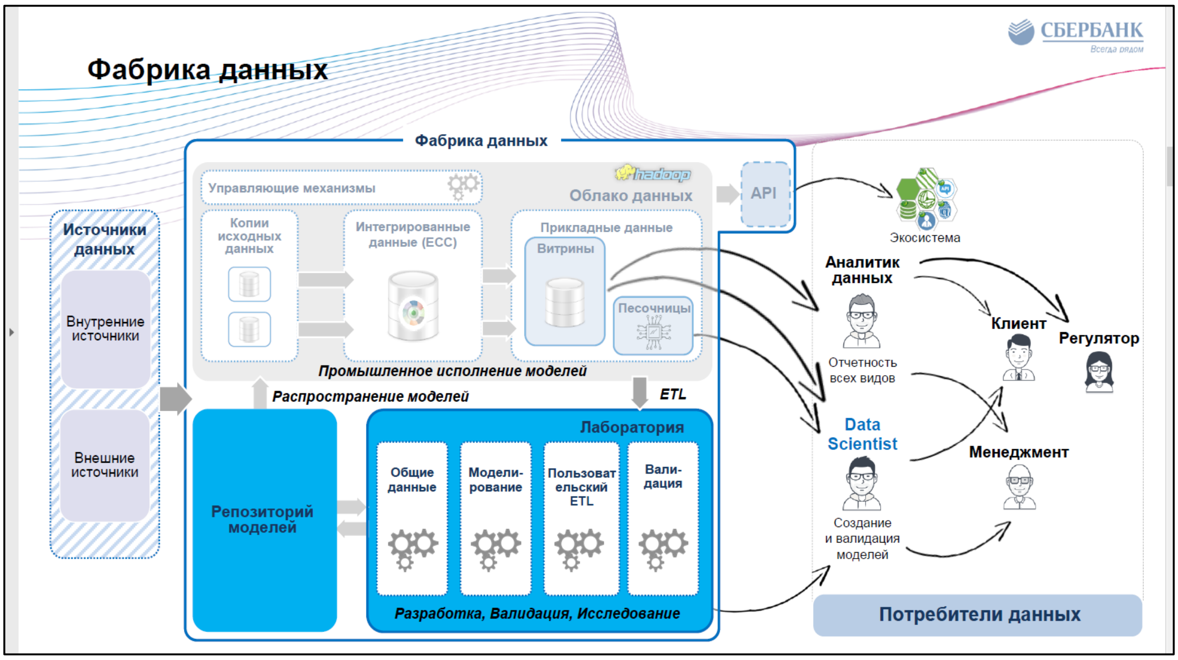
Hotels chains apply such analytics to create customized personified packages to attract new clients as well as observe the travel patterns of loyal clients to prolong relationships with them. For example, in 2017 Starwood Hotels and Resorts adopted Big Data for dynamic pricing to increase revenue. They collected external economic data, transport insights about tourist flow, social statistics to predict demand and make promotional offers and prices adjustable. For example, being a resort hotel directly depends on the weather, so when predictions are unpromising, marketing department starts planning how to attract visitor during the bad weather.

Bank sector experience

Significant contribution to Big data application in business is made by bank sector. They use Big Data benefits on a daily basis in all their activities starting from credit risk assessment till the optimisation of legal department work. Such wideness of advances analytic application requires banks to build complex IT-infrastructure. The close look will be given to Sberbank practice that is one of the leaders of AI transformation in the world

Today Sberbank is implementing a data lake called “Fabrika dannyh” where all data from internal and external systems will be stored and models will run. It consists of two parts Data Cloud and Data Lab.

The aim of the first part is to run models in production by supporting them with everyday updated data. It contains duplicates of initial data sources, their transformations, historic dashboards for modeling and dashboards for running models in production.



Pic.17 Sberbank target data infrastructure

The Data Lab was created to provide data scientists with space where they can build and tests their models not thinking about memory limits, velocity of query execution and data relevance. In the Lab every team has own space where they create necessary dashboards, do analytics researches, check hypothesis as well as write procedures and etl. What is more, the validation executed by special bank's department is carried out inside the data lab. Data Lab enables to maintain execution of following tasks :

— Data delivery

— Data preprocessing

— Data dashboards prototyping

— Pilot projects execution

— Models validation

However current architecture has several limits. For example, it is impossible to efficiently manage user spaces / resources. The process of new program/library/application installation takes a lot of time

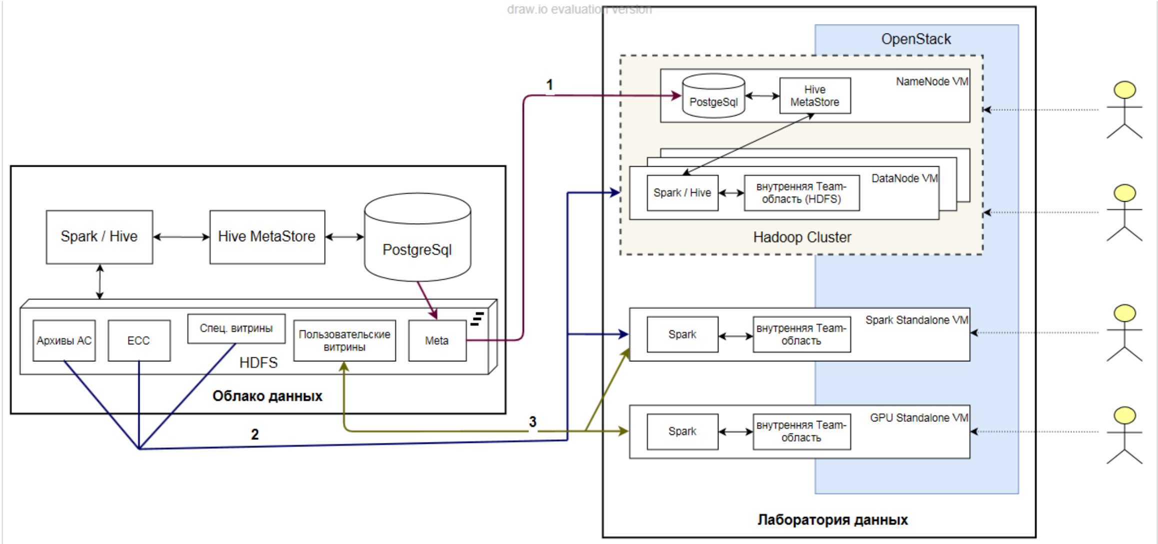
Data architectures chose stack Hadoop as a main platform for data factory because it suits to machine learning and support Python, Scala, Hive, Impala and other programming and query languages. Moreover, it allows to store and process big amounts of data. Finally, the benefits of open source solutions should be noticed.

Pic.18 IT-systems of the Sberbank Big Data adoptable platform

Naturally, as a primary data storage for application HDFS is chosen because it is a part of Hadoop ecosystem. It employs a NameNode and DataNode architecture to implement a distributed file system that provides high-performance access to data across highly scalable Hadoop clusters.

Part of the core Hadoop project, YARN is the architectural centre of Hadoop in Sberbank that allows multiple data processing engines such as interactive SQL, real-time streaming, data science and batch processing to handle data stored in a single platform, unlocking an entirely new approach to analytics. YARN is the prerequisite for Enterprise Hadoop, providing resource management and a central platform to deliver consistent operations, security, and data governance tools across Hadoop clusters. YARN also extends the power of Hadoop to incumbent and new technologies found within the data center so that they can take advantage of cost effective, linear-scale storage and processing. It provides ISVs and developers a consistent framework for writing data access applications that run iN Hadoop.

At the top level platform provides support of popular among data scientist instruments such as Scala, Python, R, Hive. They are employed to build models in data lab and test hypothesis, that is why it is level completely belong to product owners, business and data analysts. New data that daily comes to that data cloud from relational databases through very powerful ETL tool "Informatica".



Pic.19 Organisation of user access to the data

The infrastructure supports three configurations of user's connection to Data Lab : Spark Standalone, GPU Standalone and Hadoop Cluster. First two serves to process and analyse data remotely through spark ,that is why its target audience is data scientists and business analysts. Hadoop cluster configuration enables dashboard modelling and creation as well as building ETL. That is commonly used by data engineers and testers. Every access is carried out through virtual machine. Moreover, Hive of Data Lab is the only way for users to reach data in Data Cloud.

From user's point of view the space is divided into two parts: internal team-space and external team space. First is intended to save temporary users dashboards that are intermediate stages of dashboard building or data analysis. Internal team-space is available only for team that ordered a space inside VM Data Lab and only during the time it exists. External team-space is a part of Data Cloud that stores final results of modelling and data preparation. These results are available for other teams that are involved in similar projects. In fact, they are placed in the special space "Users dashboards" of Data Cloud HDFS.

Big data maturity models

For measuring the data management level of the company, the special performance indicator frameworks were proposed by IT-consulting that are called data maturity models. Models differ from one company to another but have a common idea and similar frameworks (Hortonworks, March 2016; Fern Halper, 2014). They reveal the evolution of an organization to integrate, manage, and leverage all relevant internal and external data sources. They all provide methodology to assess organisation Big Data capabilities across five domains:

- Organisation vision
- Data and analytics practices
- Technology and infrastructure

- Data management
- Organisation skills

Transformation to “data-matured” company is a journey that involves building an ecosystem that includes technologies, data management, analytics, governance, and organizational components. In order to build recommendations, companies fill in surveys where each section belongs to one of five domains. In general questions have a following content:

- **Organization vision**: to what extent does the organizational strategy, culture, leadership, and funding support a successful big data analytics program? What value does the company place on analytics?
- **Technology and infrastructure:** how advanced and coherent is the architecture in support of a big data initiative? To what extent does the infrastructure support all parts of the company and potential users? How effective is the big data management approach? What technologies are in place to support a big data initiative, and how are they integrated into the existing environment?
- **Data management:** how extensive are the variety, volume, and velocity of data used for big data analytics, and how does the company manage its big data in support of analytics? This includes data quality and processing as well as data integration and storage issues.
- **Data and analytics:** how advanced is the company in its use of big data analytics? This includes the kinds of analytics utilized and how the analytics are delivered in the organization. It also includes the skills required to make analytics happen.
- **Organisation skills**: how coherent is the company’s data governance strategy in support of its big data analytics program?

Then answers are evaluated and the company gets a final five scores. Based on survey results model says what maturity level corresponds to company data capability and consequently, recommendations are developed. The amount of levels varies from three to five where the first and the last step are almost the same but the middle stage may be as one or divided on several by the way to meet business needs.

At lower level of data maturity, a limited vision may exist solely among IT-specialists. No transformation strategy is supposed. That is why Big Data program is unbudgeted, employees are not aware of data-driven approach. Data is not collected regularly but for specific analysis related to evaluating business performance. Data warehouse is primitive. Data from different sources is not integrated. Analytics exists only in some special cases and mostly is performed manually, so no real-time analytics exists and processes are not data-driven. Concerns for security and management prevent company from exploring hybrid cloud/on-premises hosting solutions. Analytics and Development Skills are rare, usually one IT-specialist fulfils several roles.

At middle levels vision of data-driven approach is being developed, there is an understanding in top-management of the necessity of business-transformation. Data from different sources starts to be integrated in one storage however there are still departments that do not join this process. Enterprise Data Warehouse regularly receives batches of new data and analytics dashboards are build. In some cases, that can be already NoSQL or even Hadoop storage but mostly it is SQL solutions as the amount of collected data is still not really big. Business process owners request insights from an analytic department, which provides extended analytics reports. However their update still requires regular efforts. Some departments use advanced analysis to their favour. The level of security is higher than at lower level.

At highest level company has a strong vision of data-driven organisation. There is a comprehensive plan for each department and business-process about how it is ran by data. Company has a cloud storage that suits to Big Data as Apache Hadoop or Google Big Query. There all data is aggregated, transformed and analysed to be applied for real-time decision-making. Also, the historic records are stored to build machine learning models o them. Data ingested from different sources meets one standards, fully connected and does not have contradictions. Each company member understands the advantages of data-driven approach and uses it in his duty. High level of security is provided

Therefore, as soon as company’s level is found out the framework proposes the roadmap that accelerates company business-transformation into data-driven company. The models’ approach to company’s data adoption is comprehensive because it describes a complete view in terms of several departments and measures rather than points out the use of one method or technology. That is why it can be easily transferred into museum sector.

Chapter 4. Museum Big data maturity indicators development

Museum Big Data maturity levels

Having analysed other industries data practices and existing frameworks for assessment organisation data maturity, the prescriptive indicators of museum data maturity were developed. For museum purposes the five levels of data maturity are proposed with additional 0 and 5 levels that are not mentioned in the last chapter models but that suit to cultural industry realities. Moreover, the description of other three level differs a bit from business classic.

Level 0 - No data owner. At this level there is no data attention at all. Unlike private sector, museums face such situation when data analytics is not considered as essential. In this case nobody collects or analyses more data than Ministry of Culture requires. The process of digitalisation of museum collection can be started at this level. It usually happens to small and remotely located museums.

Level 1 – Data aware This level presents a state when museum staff understand the advantages of data analytics but in a limited way. A vision of systematic approach can exist only in IT-department where the interest to Big data starts growing. Though data is still not collected into one storage the potential data ingestion layer is a big because of the wide range of internal systems, media devices and online platforms introduced by museum. However, at this stage only part of these sources is considered valuable to analysing, mostly it is stored in collection management (CMS) and ticket systems. But even in this case raw data can be accessed only by vendors, whilst collection and operational managers have to use standardized limited reports available in a system interface. Data warehouse that brings together all data does still not exist. As a result, museum departments do not exchange insights between each other.

External data sources are not involved, apart from digital collection which is published online due to CMS basic functionality. There is no advanced analytics methods and consequently real-time decisions. Data security is provided at low level without serious consideration about personal data protection. Among museum staff there are no data analysts or architects, so the support of current analytics tools is carried out by IT-specialists or system vendors.

Level 2 – Data proficient. At this level, the strategy for museum Big data transformation is being developed and some practices of data management have already been implemented. Museum authority leads transformation of processes into data-driven. Long-term plans are built based on data insights. Museum process managers requests for deeper insights than inherit system modules provide on a regular basis because they know benefits. However the interest is revealed solely to related to department activities information. The data of internal sources are accessible to query by museum specialists in any time that leads to creation of data warehouse that gathers all museum raw data from internal sources and online services used by museum such as Google Analytics, Facebook Analytics, Crawlers and other instruments that calculate statistics for museum website and mobile app, Facebook and Instagram museum accounts, websites with review, museum account on audio guides websites respectively. Data warehouse is SQL database that run by PostgreSQL or Oracle DB because internal data size is not big enough to necessarily address to its storage with NoSQL database or distributed systems. No attention to external sources is given. On the whole, data infrastructure is not well prepared to share data automatically, however museum provides access to their collection through website interface, API, search engine for querying Linked Open Data and public open datasets. Due to these technologies museum shares its data with cultural knowledge aggregators and mobile applications without human intervention. Museum KPIs fulfilment are automatically calculated and sent to government database. Advanced analytics as machine learning models are used by some departments for example marketing for their own purposes regularly but not in real-time. Some tests prevented real-time analytics are held for example evaluation of building environment climate data based on object sensors measures. Privacy policy is well developed, though not the most effective state-of-the-art methods of protection are used. Museum hires as minimum two specialists to develop and supports data architecture and analytics function.

Level 3 – Data Savvy - of museum Big Data maturity implies the high level of data management that include almost complete Big Data adoption and implementation of data-driven decision making on museum daily routine. All museum staff relies on data insights that arrives in regular and visualised manner that facilitate monitoring and allow to make long-term decision while short-term are fully automated. Any high-level analytics is available via management tools without analytics skills attraction. Museum engages cloud data sources or external servers to store all data it generates and attracts. All data comes to storage meets unified standards that as a result, makes it fully associated with each other. It allows to users use museum data without regard to department it belongs. Streaming processing is employed for collection purposes such as security reason or control of artefacts sate, while visitors analytics can be processed on historic data. Machine learning algorithms are introduced in streaming processing that enables automatic decision making. External data from city public services and other cultural institutions is collected but not in real-time. Data sharing is fully automatic and does not' require human intervention. High level of data governance and security are provided.

Level 4 - Data driven - encompasses all achievements of third level and in advance enables transforming of all indicators collection processes to streaming processing including gathering data of third parties. Therefore, all connection with external sources are fully automatic and enable constant update. Batch processing stays exist only where new data comes rarely than one in hour, for example collection replenishment or ticket price change. Since fourth level is target, we look at it in details in next part.

Benchmark data architecture. The fourth level of data maturity

Target Big Data architecture

Before going to indicators let's start with describing the benchmark architecture that will define framework' main features and what it means to get high score. The view represents Big Data architecture that intends to store and analyse both real-time and historic data because only such approach to data processing allows museum to completely fulfil his missions.

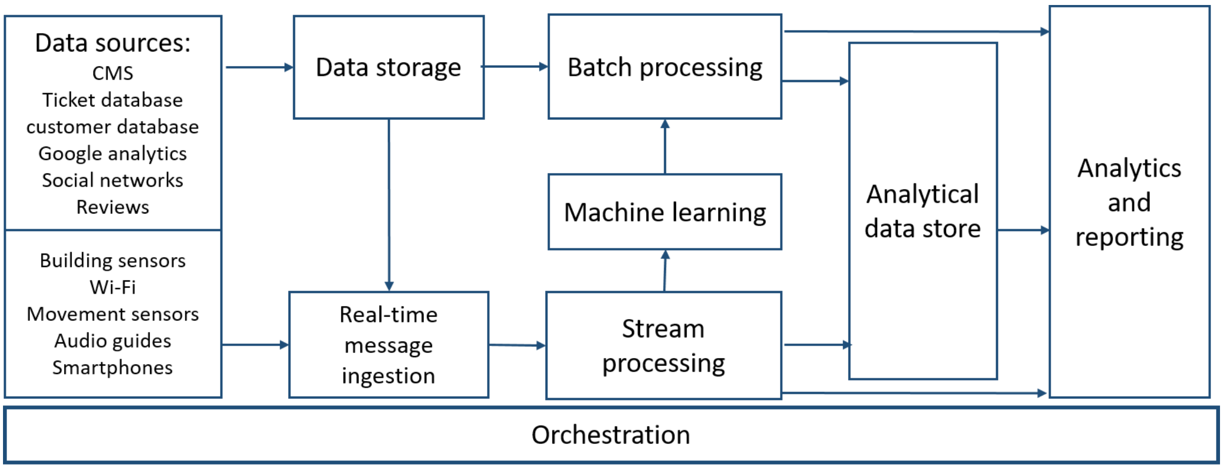


Fig.1 Target view of Big Data architecture in museum

Schematically, proposed benchmark architecture contains follows components:

- Data sources that are categorised on internal and belong to thirds parties in owner perspective, real-time and historic (batch) in terms of method of processing

Batch internal data

- Collection management system database
- Ticketing database
- Museum shop database
- Customer database (Museum Friends, news subscribers, events participants)
- Google analytics (volume of visitors does not require real-time processing)
- Social network museum accounts
- Plan business KPIs

Batch external data:

- Statistic reports of governmental institutions
- Data related to museum artefacts (literature, exhibitions etc)
- Media articles with mention of museum
- review websites
- Climate data
- City-wide statistic

Real-time internal data

- WIFI hotspots
- Visitor tracking sensors
- Object state sensors
- Building environment sensors
- Audio guides/ VR/AR
- Mobile devices
- Cameras
- Networked museum system
- Trackers in cloakroom, restroom, café, shop

Real-time external data

- City transport data including airplanes
- Other cultural organisations data

- Data Storage - Distributed data warehouse that suits to BigData and support different data formats. There all batch data is conveyed and aggregated. No data goes there does not change, thus full history is correctly presented, in order to guarantee high data quality last x dates rewrites every time. Scheduled ETL processes are responsible for writing new data. Also data storage receives pre-processed real-time data and Master data from Real-time data processing
- Batch processing – implies pre-processing of historic data . As data volumes do not allow to do ad-hocs quickly it prepares data for analysis for example by filtering or aggregating by some features
- Real-time message ingestion is necessary stage to get all signals from museum sensors in real-time and store them for short time enough to send them to stream-processing
- Stream-processing – real-time data processing that includes filtering, aggregation and other methods of data preparation
- Machine learning - group of mathematical methods that based on historic data from batch processing make predictions for new data and send results back to streaming data in order to be lately used in analytics. For example, predictions of count of visitors for next two hours based on current data and history of correlation between features like weather, day of week, exhibition topic and amount of visitors.
- Analytical data store – place where museum pre-processed historic data merges with real-time and is stored in data marts. This data warehouse serves a platform for further analytics and visualisation
- Analytics, reporting, triggering actions – the stage responded to all museum analytics. As data is already gathered to analytical data store, it can be easily and quickly retrieved to answer business request or trigger automatic actions to networked systems. It is a museum manager's instrument that varies on mission it serves to. It includes for

example changing temperature in hall or sending recommendations to stop selling tickets because of overcrowding. Among analytics opportunities OLAP cubes, BI (Business intelligence) tools as well as user reports should be named. Due to this stage every museum employee will be able to answer any his question related to museum performance

- Orchestration – the important part of data architecture that supposes maintenance and control of every stage described above. Data flows that exist between them should be attentively observed and any issue happened should be noted and fixed.

In order to be more detailed, a proposal for data architecture suitable real-time processing of data from building environment sensors is presented below. Lambda architecture of streaming processing is designed to handle massive quantities of data by taking advantage of both batch and stream-processing methods. Such approach provides balance fault-tolerance, throughput and latency.

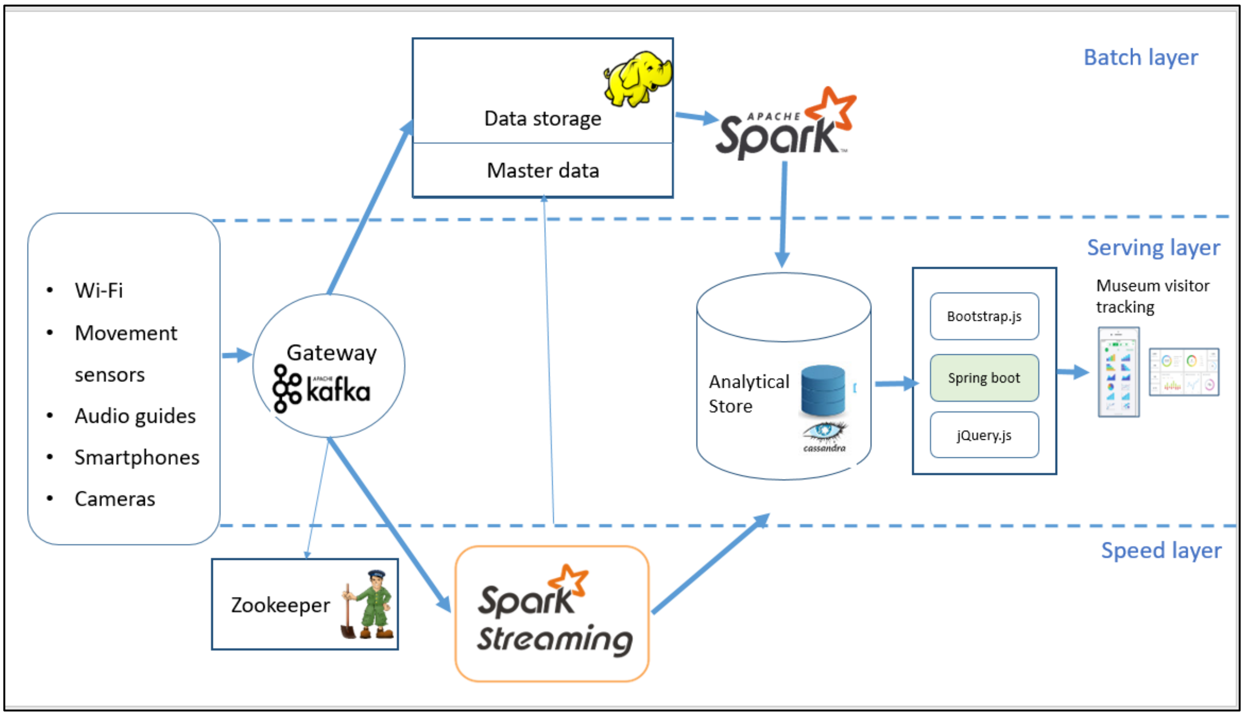


Fig.2 Lambda data-architecture for visitors tracking in real-time

From technical point of view solution with sensors looks similar to Smart Home concept. Sensors that brings data about museum environment indicators repeat functionality of Smart house sensors, part of them is networked, part presents IoT.

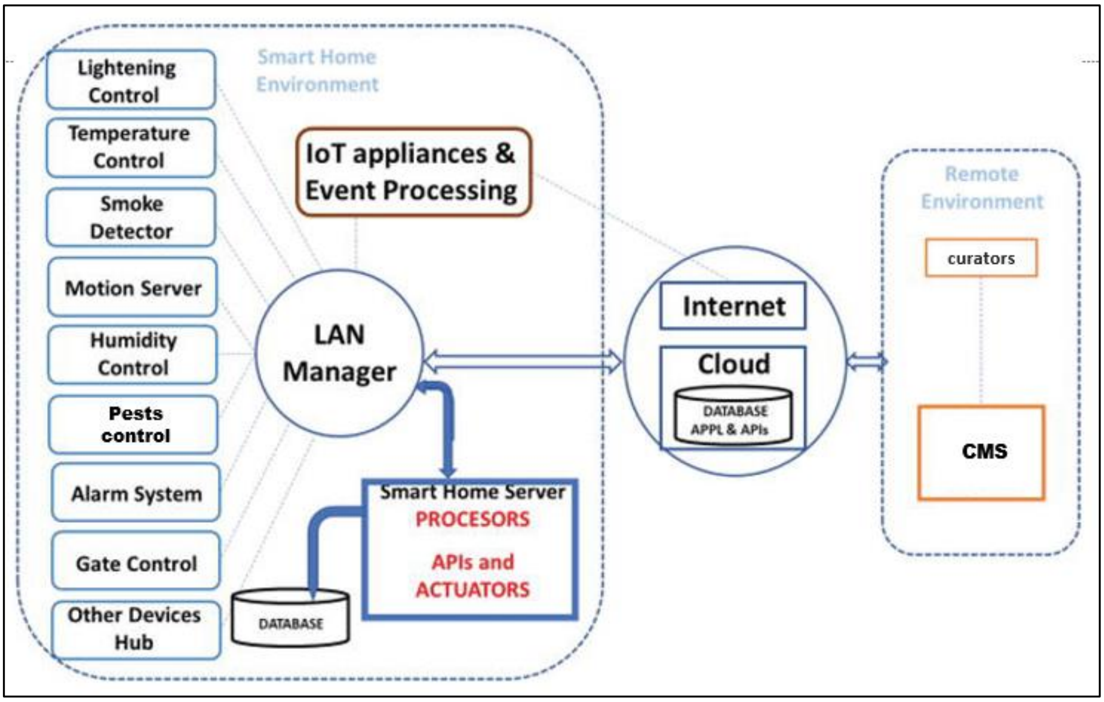


Fig.3 Target infrastructure for museum environmental control

Target data format and structure of ingested data

Museums objectives given in the first chapter dictates data properties and structure of inherit storages. Apart from common requirements as data completeness, relevance, consistency and accessibility, some of ingested data should have special features in dependence on the mission it serves.

Collection management system database:

- represents relational or graph database
- accommodates CIDOC CRM
- can be queried by SQL or SPARQL
- has a graph model at the core

Ticketing and customer databases:

- represent relational database
- provide connectivity between databases
- assign unique ID to each sale
- store unique user identifier for all databases
- store anonymize personal data

Social networks data and reviews:

- are transmitted directly to data storage
- are pre-processed with NLP methods data relates only to museum
- are anonymized data

Operational indicators:

- have format which meets governmental requirements
- are updated once in a week/ month

WI-FI streaming data:

—assign user to the ID which defined for it in the customer databases

—is anonymized data

Building analytics on data

Being data-driven organisation means making decision based on data insights coming from companies data sources in real-time. However, real-time data is a raw data and in order to adapt it to human-readable format, building analytics upon it is required. Company's management uses consolidated analytics dashboards to make decisions and create strategies. Essential factor of dashboards quality is a data connectivity in the sources used to for this dashboard. Connectivity is provided by having the same key features in different datasets from different databases. That is why it is important to look at list of **link fields** each database contains (Table 1).

Table 1. Linked fields in internal data sources

Data source	**Link fields**	**Null is permitted**
Collection database	ID_object	No
	ID_collection (group of objects)	No
	ID_exhibition	No
	ID_museum_hall	No
Ticket system	ID_transaction	No
	ID_customer	Yes
	ID_exhibition	No
	Time	No
Customer databases	ID_customer	No
	ID_customer_group	No
	ID _mailing	No
Google Analytics (browser)	ID_Session	No
	ID_customer	No
	ID_collection	Yes
	ID_exhibition	Yes
	Time	No

Media devices data	ID_device	No
	ID_location	No
	ID_customer	Yes
Movement&Object Sensors	ID_object	No
	ID_location	No
	Time	No
Building sensors	ID_museum_hall	No
	Time	No
WI-FI hotspot	ID_hall	No
	ID_customer	Yes
Smartphones	ID_customer	Yes
	ID_location	No

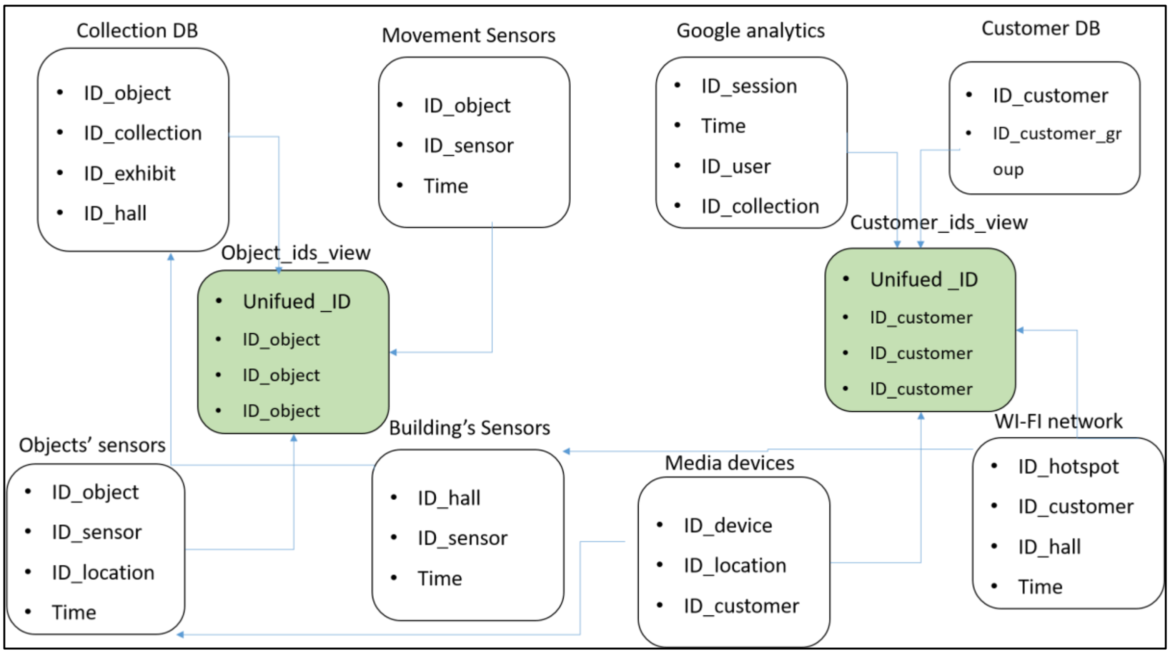


Fig4. Relations between data from different sources to enable analytics

As soon as relations between data are defined, engineers plan tables in analytic datastore and upload to there data calculated and aggregated on initial data. Such tables contain historic records which together describe the behaviour of company or company clients in the past. While sometimes statistical methods are enough to make conclusions on data, some data requires more complex technique such as machine learning models. There are different **machine learning** which can benefit museum for different reasons:

1. *Automatic recording of collection description*– algorithm of computer vision recognises digital copies of objects and automatically adds tags, short descriptions thus accelerating collection digitalisation. Also, they help attribute works of unknown authors

2. *Semantic network analysis* – algorithm works to set relations between “similar” objects of collection to simplify search

3. *Predict climate changes* – algorithm predicts museum microclimate with the aim to prevent possible negative influence on collection and reduce electricity costs

4. *Real-time prediction of attendance* – group of real-time machine learning models that based on historic patterns predict number of visitors in museums halls, such as entrance hall or cloakroom. It is used for short-term management decisions.

5. *Visitor patterns recognition* – analysis of audience behaviour based on data from wearable devices and sensors to understand exhibitions’ points of attraction, observation patterns and find out challenges visitors experience

6. *Recommendation system for newsletter content* – algorithm which says based on visitor behaviour patterns what content would be interesting for subscribers.

7. *Review analytics –method of* sentiment analysis which defines emotional colouring of reviews and picks key words that require museum staff attention.

Opening data up

Since we found out in the previous chapters that the data sharing benefits museum in knowledge enrichment and promotion perspective, it is important to include particular techniques in target data infrastructure of smart museum.

By sharing *collection* museum intends to embrace wide and diversified audience such as scientists and students, analysts, app developers and general public. Each group has own requirements to collection representation and expects museum to meet them. That is why smart museum uses following technologies:

API benefits mobile and platforms developers that aggregate collection data from group of museums in their tools as well as museum analysts that query data from collection database. API supports the following import formats: .csv, .txt, .xml, .xls, .xlsx, ASCII

Open data represents collection datasets that are available on museum website to download in the same formats as API provides. In contrast to API extracting open data does not require programing skills. The high-quality of datasets publish as open data is enabled due to hackathons and regular reuse of the data.

Linked open data is a method of publishing data based on CIDOC CRM ontology that integrates museum original metadata with other cultural sources. Within the concept museums provides a search-engine that offers researchers a flexible instrument for retrieving information about artefacts, creators, literature and events as well as associated objects by using SPARQL to get access not only to museum internal data but also to linked information in other knowledge bases. Access to link data is provided on the museum website, GitHub and international platform like Europeana that serves to integrate the cultural knowledge in one place and facilitates the information search.

VR museum and digital collection –interactive multimedia representation of museum space and its collection that passes knowledge entertaining. Data is updated once the new records are added to database.

Apart from collection data, Smart museum shares its operational and attendance statistic with external organisations. First of all, they send it to the Ministry of Culture and city council in order to provide authorities with necessary statistic. Moreover, their data with data of other art institutions or public services lay at the core of Smart City concept.

Legislation

Working with data, especially personal one, demands a high attention to legal acts that regulate methods of data collection and ways of its usage. For museum there are two acts to consider foremost – Personal data low (European analogue The General Data Protection Regulation) and Common creative licence (CC0).

Personal data low regulates museum privacy policy that sets rules for personal data processing. Any collection of personal details such as name, gender, date of birth, email etc. that museum receive should be approved by a person and accompanied by detailed privacy policy. Situations when museum gets personal data are the next:

- visitor connects to WI-FI network
- apply for membership
- purchase online tickets,
- subscribe to newsletter
- sign up in mobile app
- register to event
- contact with museum

The use of personal data is a very sensible topic that makes museum carefully store and publish visitor's data. Among possible purposes are marketing, general use, administration, internal research and profiling, security. Data cannot be shared with third parties except selected partners that provide museum with services, for example payment processors.

According to the law, museum must fulfil every person request to get the list of sources where his data is stored and how museum uses it. Data that enters to museum database should be depersonalized by data masking algorithms such as substitution, shuffling or encryption. That means no one from museum staff cannot see the original personal information.

Common creative license (CC0) has an opposite influence on data sharing. CC0 is a public copyright license that enables a free distribution of published content without any restriction of use, copy or distribution of data in any other resources for any purpose. Collection's API and search engine release, open datasets and digital image publication are available to use only under CC0 license.

Human resources

Modern museum practices prove the advantage of having IT-specialists with programming and analytics skills, no matter whether they accommodate end- product or develop their own from scratch. At least, one mobile developer that is experienced in creating API, one data architect with Big Data knowledge and two data analysts should support museum data ecosystem.

Process management

However, not only data analysts but all museum staff work with data and have an awareness of ongoing data-transformation and reasons for it. Business processes are fully automated, so curators and managers only monitor their efficiency and develop strategies based on current performance. Museum management makes decisions solely based on data-insights.

Museum Big Data maturity Indicators. Version 1.0

Table 2. Data ownership

Indicator name	Question	Business function	Object to evaluate	SCORE				
				1	2	3	4	0
				Source does not exist at all	Source exists but museum staff does not have an access to it	Limited access: only some reports are available	Unlimited access of museum staff	Source does not exist for reasons beyond museum control
Data accessibility	How easy is to access data from this source	Collection management	Collection management system database					
			Data from collection objects sensors					
			Data from building sensors					
		Visitors management	Ticket system database					
			Museum shop database					

			Audio guides data					
			Smartphones data					
			Bahavior data from VR/AR devices					
			Data from wi-fi network					
			Movement data from sensors,cameras , WIFI					
		Customer relationships/ Social engagement	Museum friend database					
			Google analytics data					
			Data from social networks accounts					
			Loyal audience database (subscribers, event visitors etc)					
		Operational management	Target KPIs dataset					
			Financial database					
Data quality	What weeknesses does data have in sources			**Source does not exist at all**	**Low quality**	**Average data quality, some problems exist**	**High data quality, no problems with data**	**Source is not available for reasons beyond museum control**
		Collection management	Collection management system database					
			Data from collection objects sensors					
			Data from building sensors					
		Visitors management	Ticket system database					
			Museum shop database					
			Audio guides data					
			Smartphones data					
			Bahavior data from VR/AR devices					
			Data from wi-fi network					
			Movement data from sensors,cameras , WIFI					
		Customer relationships/ Social engagement	Museum friend database					
			Google analytics data					
			Data from social networks accounts					
			Loyal audience database (subscribers, event visitors etc)					
		Operational	Target KPIs dataset					

		management	Financial database					
Data structureness	How does data structure suit to machine-readable format?			**Source does not exist at all**	**Completely unstructured data**	**Some pieces require transformations to get structured**	**Presented in machine-readable format**	**Source is not available for reasons beyond museum control**
		Collection management	Collection management system database					
			Data from collection objects sensors					
			Data from building sensors					
		Visitors management	Ticket system database					
			Museum shop database					
			Audio guides data					
			Smartphones data					
			Bahavior data from VR/AR devices					
			Data from wi-fi network					
			Movement data from sensors,cameras , WIFI					
		Customer relationships/ Social engagement	Museum friend database					
			Google analytics data					
			Data from social networks accounts					
			Loyal audience database (subscribers, event visitors etc)					
		Operational management	Target KPIs dataset					
			Financial database					
Data coherency	What part of internal data sources can be linked with each other			**All data sources are incoherent**	**Some data sources can be coherent after transformation**	**Data inside one business function is coherent**	**All data can be linked without transformation**	**Source does not exists or is not available for reasons beyond museum control**
		All	**All**					
Data availability	At which manner does museum acces to the source?			**No information, source has never been accessed**	**Manual not regular access**	**Regular manual access**	**Automatic scheduled access**	**Source does not exists or is not available for reasons beyond museum control**
		Collection management	Government database of cultural heritage					
			Collection linked open data of other institutions					

			Published open data sets related to collection					
			Online text sources such as scientific journals and reports, conference materials					
		Visitors manageme nt	Climate data					
			City services open data					
			Mobile apps/Audioguid e apps					
			Tourist data from tour operators, hotels, air planes					
			Third parties relevant data (f.e. mobile operators , search engine)					
			Data from websites with reviews					
		Customer relationshi ps/social engageme nt	Government institute statistic					
			Other NGO open data					
		Operationa l manageme nt	Ministry of Culture targets					
Data structure ness	How does museum stores data?			**Received data is not stored**	**Data is stored in text unstructured view**	**Some fields are redundant and should be divided**	**Store in structured view**	**Integrity does not exists for reasons beyond museum control**
		Collection manageme nt	Government database of cultural heritage					
			Collection linked open data of other institutions					
			Published open data sets related to collection					
			Online text sources such as scientific journals and reports, conference materials					
		Visitors manageme nt	Climate data					
			City services open data					
			Mobile apps/Audioguid e apps					
			Tourist data from tour operators, hotels, air planes					

			Third parties relevant data (f.e. mobile operators , search engine)					
			Data from websites with reviews					
		Customer relationships/social engagement	Government institute statistic					
			Other NGO open data					
		Operational management	Ministry of Culture targets					
Data integrity	What part of external data sources can be linked with internal sources			**Data fom any source cannot be integrated with internal**	**Some sources can be linked auto transformation**	**The part of sources has integration with museum data**	**All sources are possibly integrated with internal sources**	**Source does not exists or is not available for reasons beyond museum control**
		All	**All**					

Table 3. Data collection&processing

Indicator name	Question	Business function	SCORE					
			1	**2**	**3**	**4**	**5**	**0**
Data depth	**When is the oldest data related to business function availble outside internal source? (Several answers possible)**		**No collection is suggested**	**< 1 month ago**	**<1 year ago**	**1-3 year ago**	**>3 year ago**	**Collection is not available for reasons beyond museum control**
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operational management						
Data storage	**Where is data ingested? (Several answers possible)**		No collection is suggested	In files/documents	In siloed databases across the organization	In central data warehouse (SQL)	In Big Data optimized storage	Collection is not available for reasons beyond museum control
		Collection management						
		Visitor management						

		Customer relationships/ Social engagement						
		Operational management						
Data granularity	**What is the most detailed information available? (Several answers possible)**		No collection is suggested	Final indicator (e.g. amount of exhibition visitors)	Information aggregated by one measure (e.g. amount of visits of one customer)	Inherit data with unique key field (e.g. customer, ticket number, enter_ID..)	In case of real-time collection raw data as is	Collection is not available for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						
Ingested source variety	**What part of available source does participate in data ingestion?**		No collection is suggested	<=10 %	<=50%	<=75%	>=75%	Collection is not available for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						
Data update	**How often is data conveyed ? (Several answers possible)**		No collection is suggested	By request	Periodically but not often then one in month	Regularly according to schedule (e.g. daily)	In real-tme (approx one in 30 min)	Collection is not available for reasons beyond museum control

		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						
Data governance	**What is the level of data governance?**		No collection is suggested	ad hoc governance structures are in place	Basic data governance is in place to help ensure data quality and security	Advanced data governance is in place to ensure quality, security, and correct access over the complete life-cycle of the data	Data governance ensures correct data management over the complete data life-cycle	Collection is not available for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						
Data exchangability / connectivity	**What level of connectivity does collected data provide ? (Several answers possible)**		No collection is suggested, so data can be exchanged	Data connectivity is provided within one department	Data connectivity is provided among several department	Full data connectivity is provided within museum	Exchange with external sources	Collection is not available for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						

Table 4. Data analytics

Indicator name	Question	Business function	SCORE					
			1	2	3	4	5	0
Sophistication of methods	What of the following methods does museum apply? (Several answers possible)		No analytics exists	Manual (Excel)	Ad-hoc (answer a single, specific business question) - queries to DW	Relevant BI Dashboards	Machine learning	Analytics is impossible for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operational management						
Depth of analysed data	What maximum depth of data is extracted for analysis? (Several answers possible)		No analytics exists	Last reporting period (day, week, month)	Last several reporting periods	All available history	Streaming processing/ Real-time	Analytics is impossible for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operational management						
Wideness of analysed data	How many sources are embraced for analysis? (Several answers possible)		No analytics exists	Only own source	All sources available in department	All museum available data	External data	Analytics is impossible for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operational management						
Data influence	How do data insights influence museum processes? (Several answers possible)		No analytics exists, i.e no influence	No influence, analytics is used for reporting purposes	Analytic insights are used by some departments for short-term plans	Analytics insights influence all museum processes and common vision	Analytics is used for real-time decision-making	Analytics is impossible for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						

Level of analysis depth	What is the most detailed information available? (Several answers possible)		No analytics exists	Final indicators (e.g. amount of exhibition visitors)	Information aggregated by one measure (e.g. amount of visits of one customer)	drill down reports (allows to navigate to a different layer of data granularity)	Machine learning predictions	Analytics is impossible for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						
Analytics customers	Who benefits from data analytics insights? (Several answers possible)		No analytics exists	Employees responsible for analytics	Museum management	All Museum staff	external environment	Analytics is impossible for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						

Table 5. Technologies&Infrastructure

Indicator name	**Question**	**Business function**	**SCORE**					
			1	**2**	**3**	**4**	**5**	**0**
Data infrastructure level	What type of data infrastructure does support analytics related to business function?		No infrastructure exists to support analytical activities.	Integral program modules are used to support analytical activities	There is standard analytic infrastructure that supports all data processing stages	Distributed systems or cloud services that supports Big Data are introduced	Architecture supports streaming-processing	Infrastructure does not exist for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						
Capabilities	What skills do exist in museum to enable infrastructure performance		No infrastructure exists to support analytical activities.	All analytic tools are supported either by vendors or IT-specialists	There is a senior data achitect that is skilled to create and support full data processing	There is a data architect that is skilled in Big Data infrastructure creation with junior analyst	There is a team skilled in Big Data infrastructure creation	Infrastructure does not exist for reasons beyond museum control

		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						
Sophistication of data exchange	What instruments does museum support to share data		No infrastructure exists to support analytical activities.	Manual file upload to resource	Automatic file upload to resource	Providing automatic access to data source (API)	Real-time access	Infrastructure does not exist for reasons beyond museum control
		Collection management						
		Visitor management						
		Customer relationships/ Social engagement						
		Operation management						

Table 6. Museum readiness

Indicator name	Question	SCORE				
		1	2	3	4	5
Smart museum vision	Who has a vision of Smart Museum?	No vison exists	Vision exists only in IT-department	Museum management realizes potential of data-driven aproach	Vision of data-driven approach in all museum	Vision of smart museum concept at every level of museum
Analytics focus	What question does current analytic mostly answer?	Analytical effort was not started	Focus is on describing; what happened?	Focus is on diagnosing; why did it happen?	Focus is on predicting: what will happen?	focus is on prescribing; how will we make it happen?
Museum staff attitude	How are museum employees aware of the possibilities of analytic?	Unaware at all	Skeptical of the benefits of a data-driven approach	Museum Staff involved in statistic and monitoring are interested in using analytics	Employees are engaged in getting the most out of the data and analytics.	Actively promote analytics to other cultural institutions
Role of analysis	How are museum employees aware of the possibilities of analytic?	Analytics is not featured in the strategy of the organization.	Analytics has an ad hoc and experimental place in the strategy of the organization	Analytics and data-driven adoption have an official place in the strategy of the organization	Empowering all employees to get the most benefit out of data and analytics is the major focus of the strategy.	Analytics is the major driving force of museum

Data literacy	How are museum employees ready to deal with data?	No analytic skills among employees	Only people that are responsible for statistic are skilled	All museum employees have basic data literacy skills	All museum employees have statistical knowledge they use to set up experiments	All museum employees in the organization have the knowledge to innovate in new products and services using data and advanced analytics

Chapter 5. Framework application in Russian museums

Assessment of the Pushkin museum of Fine Art data maturity

The advantage of developed approach to Big Data maturity assessment is its easy distribution on museums. In this paper, the model was tested on the Pushkin museum of Fine Arts that introduced the one of the best approaches to the adoption of digital innovation to museum routine.

The framework was filled in based on author's data and technical infrastructure audit and museum staff survey results. Observation of current infrastructure allowed to mark the first four group of indicators: Data ownership, Data collection &processing, Data analytics and Technology & Infrastructure. With help of IT-department Museum readiness was also assessed.

The model showed the Pushkin state museum Big data maturity level is **Data proficient(2 level)** where Data Ownership maturity is scored to the second level and Museum readiness to the third. That is a good sing because it proves the museum's high potential.

Table 7 The Pushkin museum level of Data ownership

		1	2	3	4	0
Total:	Collection management	3	1	5	0	0
	Visitors management	1	2	1	0	0
	Customer relationships/ Social engagement	0	2	4	0	0
	Operation management	0	4	0	0	0
	Level of data ownership	**3**				

Table 8 The Pushkin museum level of Data processing

		1	2	3	4	5
Total:	Collection management	0	2	0	0	0
	Visitor management	0	2	0	0	0
	Customer relationships/ Social engagement	1	0	1	0	0
	Operation management	0	1	1	0	0
	Level of Data collection&Processing	**2**				

Table 9 The Pushkin museum level of Data analytics

1	2	3	4	5

Total:	Collection management	0	2	0	0	0
	Visitor management	0	2	0	0	0
	Customer relationships/ Social engagement	1	1	0	0	0
	Operation management	0	2	0	0	0
	Level of Analytics	**2**				

Table 10 The Pushkin museum level of Technologies

		1	2	3	4	5
Total:	Collection management	0	1	2	0	0
	Visitor management	0	1	2	0	0
	Customer relationships/ Social engagement	0	2	1	0	0
	Operation management	0	2	1	0	0
	Level of Technology	**2**				

Table 11 The Pushkin museum level of Museum readiness

Total		0	2	3	0	0
Level of Museum readiness	**3**					

Due to the model the next *challenges and point for development* were revealed:

Data ownership:

— External sources are not addressed at all

— No sources of real-time analytics are used

— There are issues with quality of internal data that leads to week connectivity

— Data is not prepared to be integrated with external resources

— The limited access to ticket and customer database does not allow to build analytics

Data collection &processing:

— Analytics data architecture is not suggested in museum

— Data is stored chaotically, only some departments has historical data

— Apart from WIFI outputs, history is kept in documents

— The low level of data granularity, rarely detailed information is available

— Data is ingested from less than third of owned sources

Data analytics:

— Analytics exists only for reporting purposes

— There is no regulated approach to analysis

- Only three business function benefit from analysis
- Small share of data is analysed
- Analytics methods are primitive
- Analytics influences very limited part of business processes
- Small part of employees use analytics insights

Technology&Infrastructure:

- No special data infrastructure, reports are supported by systems integral modules
- Data architect skills are not supposed
- Manual data exchange with third parties

Museum readiness

- Smart museum vision has already started to form, however only IT-department is aware of its details
- Museum staff does not know why they need analytics

Recommendation for the Pushkin museum of Fine Art

Based on received results, recommendations were formulated. They seek to help museum to reach the Data-Proficient level and build the strategy of museum transformation into data-driven organisation

Concerning **collection management,** the main improvements should be made with aim to make the data use more comfortable as for museum staff as for external consumer.

For museum own aims system database should be available for analytics. It means data is ingested by scripts (SQL usually) and sent to museum data storage. Technically, such change takes no more effort than asking for the vendor confirmation.

First, to make data sharing possible the licence status of the content should be determined. *Creative Common licence* is a choice of the most organisations which share data for copyright disclaimer.

Next, the museum database collection should be available online. It requires the relook of all collection data storage. Since Kamis provides a relational database model, however, it does not meet standards for sharing. Graph model defines relations between objects in the same way as data is saved in the Internet. That is why it allows search engines to find them and their metadata. The best way for the museum is to incorporate *CIDOC CRM* standard for database. In this case (Pic. 20), data is still stored in relational database, but there are also tables of relations and tables of objects which are necessary for storing Linked data.

Objects

ID	NAME_CIDOC	NAME2
177900	E28	Conceptual_Object
177910	E38	Image
177893	E20	Biological_Object
177878	E5	Event
177892	E19	Physical_Object
177880	E7	Activity
177897	E25	Man-Made_Feature

Relations

ID	NAME_CIDOC	NAME2	NAME3	SCOPE_DEF
178066	P67F	refers_to	(null)	(CLOB) Thi...
178229	P53F	has_former_or_current_location	(null)	(CLOB) Thi...
178051	P39F	measured	(null)	(CLOB) Thi...
178104	P14F	carried_out_by	(null)	(CLOB) Thi...
178228	P74F	has_current_or_former_residence	(null)	(CLOB) Thi...
178160	P46F	is_composed_of	(null)	(CLOB) Thi...
178053	P62F	depicts	(null)	(CLOB) Thi...
178122	P16B	was_used_for	(null)	(null)
178265	P16F	used_specific_object	(null)	(CLOB) Thi...
178082	P44F	has_condition	(null)	(CLOB) Thi...
178083	P5B	forms_part_of	(null)	(null)

Pic. 20 CIDOC CRM adoptable database

While the graph data model benefits researchers, the museum should also think about opening data to developers. Consequently the next step is the release of collection API that is easily created and supported with data Graph model. An essential part of our planning process is to consider how museum content might be most useful to our audiences.

Here, there are some questions important to consider before starting API development:

- How suitable is the writing style of the content?
- How suitable is the level of technical complexity of the content?
- Will it suit to potential users?
- Are there staff in your organisation who will be able to re-write the thousands to millions of these pieces of content?
- If so, who should rewrites be tailored for - the researcher, the tertiary or post-graduate student, the school pupil, the pupil's teacher etc?

When the questions are answered, museum can start API development. Museum can either develop it from scratch or use turnkey solutions with risk of bugs fix delay. In case of Pushkin Museum, it possesses enough resources to create its own API. The possible realisation is to use Linux platform for serving the API, write code on Python/Django. During the design, consideration should be given to enabling the possibility of moving the site out into Google App Engine. Apache runs the scripted pages and nginx serves the static content. The proposed database is PostgreSQL.

Another important thing about API is that Museum should use the tool also for its own needs. It is the only case to guarantee the high quality of API.

At the whole, the data infrastructure for collection management is presented on Figure 3.

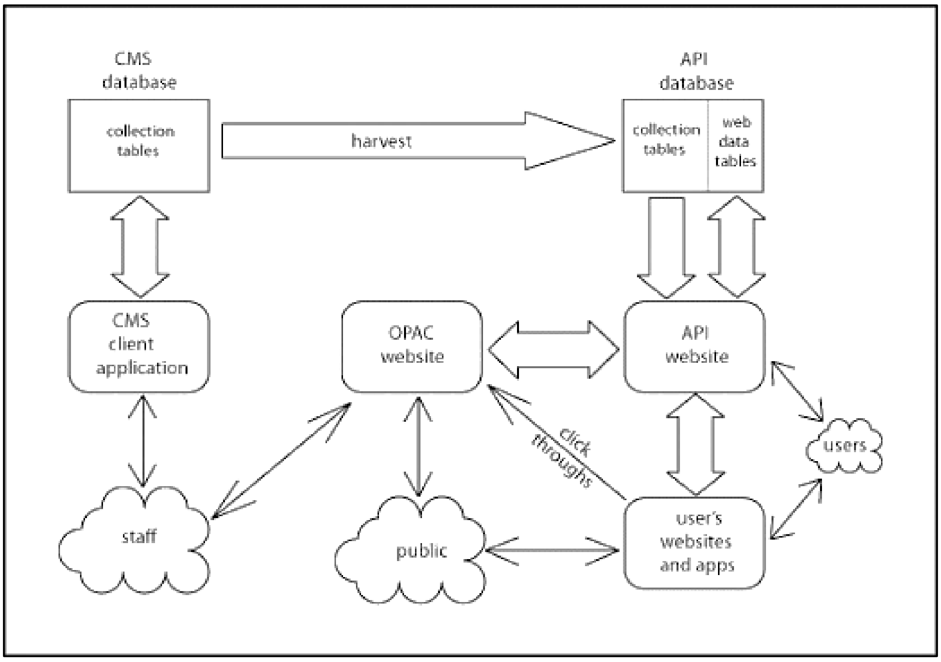


Fig. 3 Architecture that enables efficient collection sharing

As for visitors and customer management, the museum needs to get an access to exploited data from all systems and ensure its consistency. To do this, It is important to we discuss all data sources separately.

First, data from WNAM has already stored in MongoDB on Museum local server and the remote connection is set. But as MongoDB is NOSQL database, it is not suitable for querying data, hence the Museum needs to choose alternative database where queries will be easily processed. The second concern is about limited visitors information accessible to the museum. The system potentially is able to collect more data about visitors. WNAM functionality includes authorization by email or social network account option which allows to identify big amount of users and track them by WIFI signal. Also, an option to subscribe to museum news when connecting to WIFI proposes an opportunity to increase the number of museum followers. In total, following these recommendations museum can collect the users database which significantly increase the level of connectivity between its data sources.

Regarding CRM system, first of all, the museum should assure data consistency. Database must not have gaps in key fields because it makes impossible to identify unique records and link tables. The simplest way for museum to guarantee this is to make these fields obligatory. Furthermore, museum should extend the list of obligatory fields by adding such fields as «email», «phone number», «gender» and question «how you know about "program". Also, the museum should set remote access to CRM database that is located in a local server and create a special user account to have possibility to regularly extract data from it.

Next, today the museum does not have a read access to TicketNet database which is necessary to build analytics reports. Despite it is located on Museum servers and technically there are no challenges to extract data from the database, legally only provider has a right to access. However insights that can be gained from this database are valuable for visitors analysis. For example, they comprise museum "Friends" passages, offline and online ticket sales statistic and level of attendance per day in different buildings. Accessing this data would enable museum with detailed information about 'friends' behaviour and popularity of held exhibitions.

Google analytics provides the essential information about visitors online behaviour but it also requires modifications. As it was noticed above, the event tracking should be set for various events, such as clicks on exhibition labels or clicks on "Fridays in Pushkinskiy" sign or purchase of Museum Friend card. It will enable Museum to find out visitor's interests. Search analytics report is also an important thing that can be added. It helps a museum to collect texts of search queries made by users coming to its website. Due to this, it can find out their motivation to open the museum website. Museum also can consider about adding context advertising and AdWords as instrument to manage them. We will use Google Analytics API to download data from it.

As for Facebook and TimePad, museum can also use their APIs that allows automatic and safe access to data they store. Moreover, it would be recommended to add GA Client's ID to TimePad, thus tracking and collecting data about visitors of TimePad published events.

Moreover, in the current infrastructure there is no place where subscribers and other loyal clients data will be gathered systematically.

Below there is a view (Fig.4) of how data from different sources should be connected to enable relations in analytics data storage . It will make visitor analytics possible. Yellow colour means that there are some challenges with data

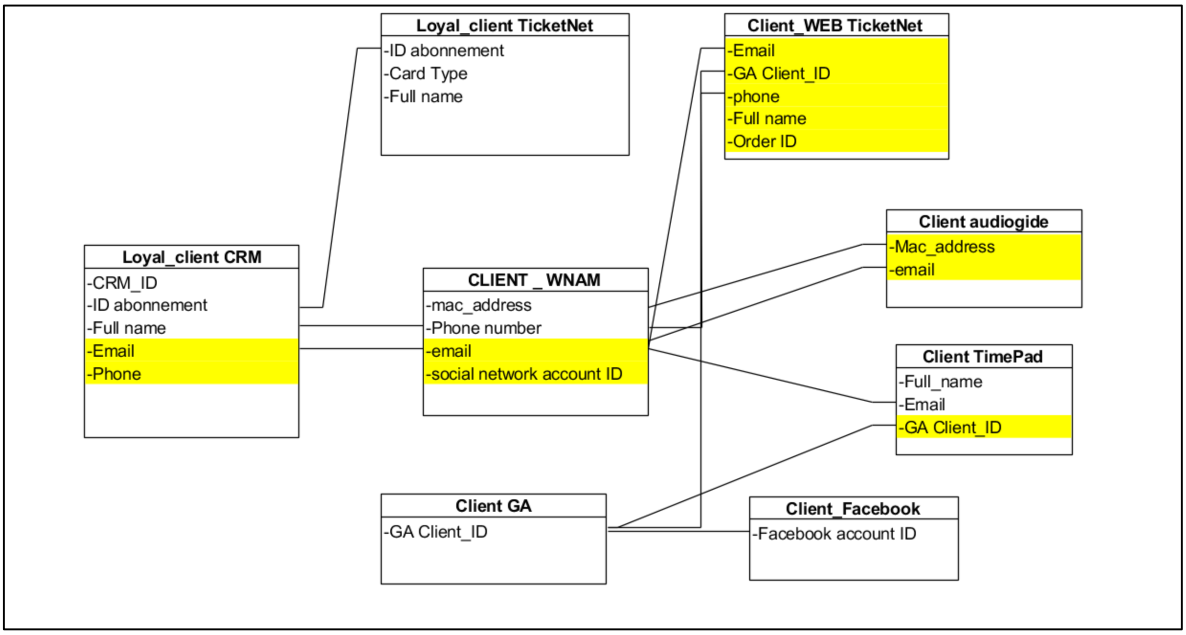


Fig4. Visitor data target connectivity

Now when we suggest that all sources have good data quality and structure, we should consider such data architecture that will allow (1) to connect data from all systems, (2) regularly update it, (3) store historical data, (4) provide easy access on-demand for different users, (5) be reliable and guarantee safety of information, and (6), finally, let business users easily view reports.

The architecture with data warehouse in the centre would meet the criteria of Data Proficiency level. There ETL processes regularly transmit data from parent system's databases to analytical store. With the data aggregated in this store BI tool visualizes dashboards.

Taking into account the current medium amount of data belong to museum, PostgreSQL is the best option for data warehouse in short-term perspective. Although it is not target decision. Lately, Postgres will be replaced by a highly scalable warehouse as Yandex Click House that can be placed in Yandex Cloud or by Hadoop that will enable museum to implement real-streaming.

It is a powerful, open source object-relational database system with over 30 years of active development that has earned it a strong reputation for reliability, feature robustness, and performance. PostgreSQL offers enterprise-grade features including a strong emphasis on extensibility and standards compliance. PostgreSQL runs on all major operating systems, including Linux, Unix, and Windows. It is fully ACID-compliant, and has full support for foreign keys, joins, views, triggers, and stored procedures (in multiple languages). So it is open-source relational database that is easy installed and maintained.

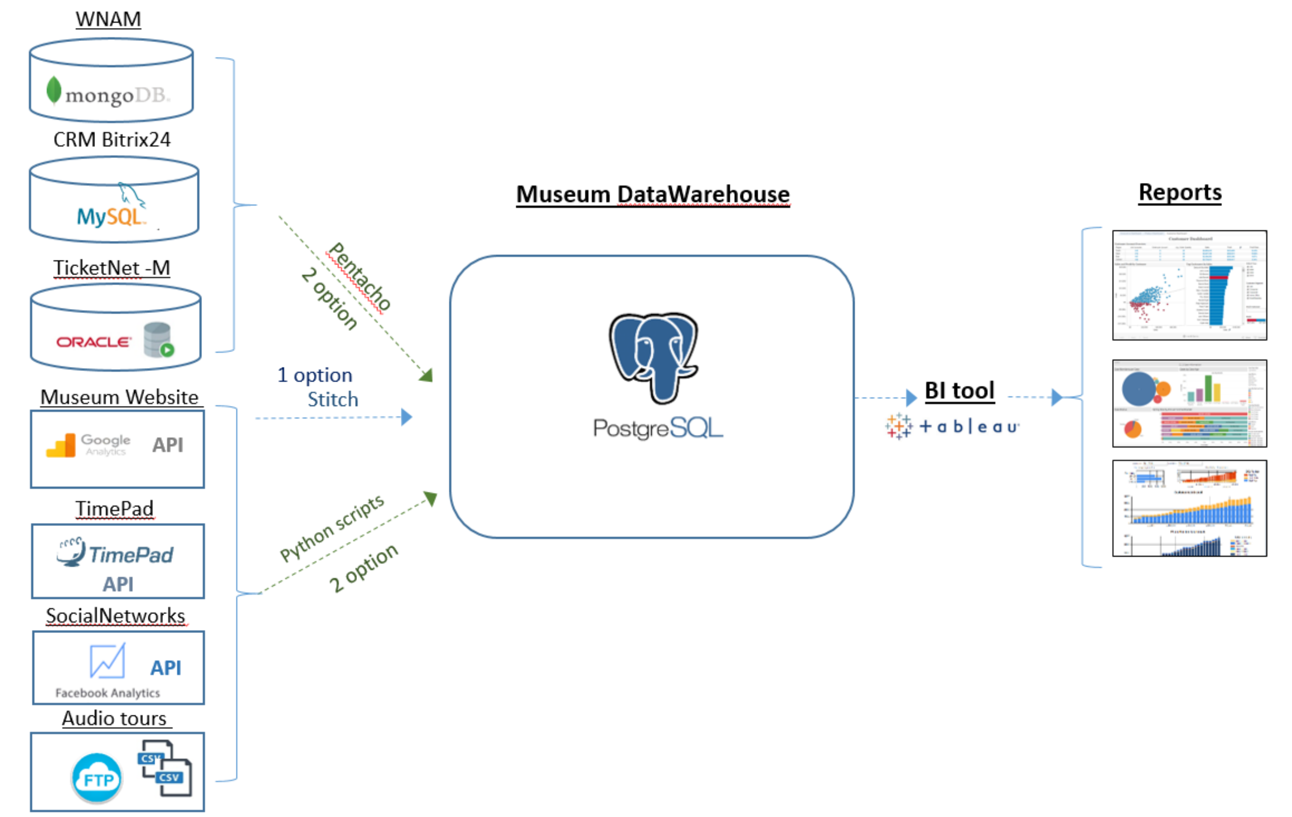


Fig 5 Scheme of Level 2 Data Infrastructure

In that architecture ETL (Extract, Transform and Load) processes are responsible for extracting data from source systems and bringing it into the data warehouse. What tool should be chosen for building ETL depends on Museum budget and human resources. There are two options: first is exploit free tool Pentaho Data Integration or Kettle for extracting data from databases and use Python scripts to get data from external sources and Google Analytics. In addition all ETL jobs should be croned means to scheduled.

Actually, Kettle is a component of Pentaho but can use separately. It does not demand special skills except basic SQL knowledge as every process is created with a graphical tool where user specify what to do either without writing code if there are no conditions or with writing one if user wants to define some conditions; because of this. However, to build all these pipelines, test their designs to ensure they runs smoothly and what the most critical to maintain them that means catch errors, recover, edit and so on museum will have to hire an additional specialist. Besides, museum needs an employee maybe the same that has skills to write Python code to bring data from Google Analytics, Facebook and TimePad APIs to the warehouse.

The other option is to buy ETL system like Stitch that provides completed data pipelines for the most famous databases including MongoDB, MySQL and soon Oracle. It allows to collect events directly within minutes without API maintenance from the products as Google Analytics or Facebook Ads. Stitch has a rather primitive interface and allows user to simply tick any tables and any fields he wants to replicate, with what frequency and in what way. Stitch supports full overwriting and incremental key replication, which can be, for example, a primary key if the rows in the source table are only added but not changed. Also it monitors the structure of the source database and changes in the tables — new columns are picked up automatically. Moreover, Stitch is cheapest among its analog and gives a free trial for a month, during which you can upload all the historical data and set up incremental replication of new data. As for price there are four tariffs that depends on volume of processed data. For Museum volumes cheapest tariff plan “Starter” that imply 5 mln transformed rows per month is sufficient. It costs 1000$ per years that’s about 60000 rub. It is much cheaper than hire one more employee.

Selection of BI system is difficult question and depends on exact business requirements and demanded frequency of reports. However, the best one in that area is Tableau. User does not need any knowledge in visualization to use it as it has native interface. It queries relational databases and present tables fields as parts of constructor so you can build any report just by moving them to need field. Moreover, in last versions they provide base machine learning tools that allows to build predictive models without hiring data scientists.

Now when we have the model of museum infrastructure and it will allow us to create portrait of museum visitors just tracking him in different source. To do it we need to connect tables by client’s keys that we extract from WNAM system as “phone number”, “email”, “mac-address”. When all tables are in one data warehouse it will be easy to link them with each other and build dashboard based on it.

As example, let’s examine a task how to find out the portrait of museum loyal visitor. Business would like to know what people are likely to be interested to join loyalty program. In order to solve this problem, we can build classification predictive model that will say us based on the client’s portrait whether he tends to join “Museum Friends” program. Then we will offer him to join this program via email newsletter or while he is connecting to Museum WI-FI network. As a result, Museum will get more loyal clients and extra profit as well. Certainly, such model requires qualitative data that will be provided by proposed data infrastructure.

Therefore, the short-term museum task is to enable access to all possessed data sources, enable the acceptable quality of data and its connectivity.

Certainly, it requires analytics skills inside museum. In order to achieve the second level of Big data maturity senior data architect and data analyst are essential, at least to support data ecosystem and provide analytics reports

Lastly, it is crucial for museum to build comprehensive strategy and make smart museum vision is one of key targets. Therefore, people close to analytics as well as top management will understand significance of data-driven approach achievement which is a minimal requirement.

Conclusion

Summing up the research findings, there is a big range of possibilities data application in museums. All cases given in the work prove the advantages of using data-driven decisions in all museum processes, starting from collection management to social contribution. Therefore, museums should take into consideration the development of data culture and move toward the fifth level of Data maturity described in the elaborated framework. Only this level allows museum to seamlessly be integrated into the Smart City strategy and named Smart Museum.

However, the indicators goal is not limited by museum needs, it also aims to benefit Russian government and, especially, the Ministry of culture. Museum Big Data maturity model is an effective instrument that allows authorities to get relevant and objective view on Russian museums current state. It highlights their strengthens and weaknesses in terms of digital transformation. Aggregating indicators by region or other measure will show statistics that will help to better develop Russian smartness.

The other motivation for the government to help museums implement ‘Smart vision’ is an opportunity to increase the efficiency their controlling function. Due to digital transformation all data that now museums are collection and sending to the Ministry manually will be automated, thus decreasing human factor. As a result this will improve the quality of transferred data, its structure, coherence, and depth. According author’s believe it should accelerate Russian cultural industry growing.

The proposed by this work indicators framework is a first approach to assess museum data maturity that should and will be broaden and deepen in the future. The easy distribution on all Russian museum industry enables museum IT-specialists to define their museums level of data maturity, recognise problems and point to improvement similar to the work done in the Pushkin museum of Fine Arts. Application of the model in other museums will allow to improve model, add more details or consider some analytics possibilities or in opposite challenges that, finally, make it ready to become a part of Smart Museum Index.

Bibliography

(Editor), J. M. (2013). *New Museum Theory and Practice: An Introduction.* Blackwell Publishing.

Agency of strategic Initiatives. (Juny 2018). *Приоритетные направления внедрения технологий умного города в российских городах.* Moscow.

Alexander, E. P. (1995). *Museum Masters: Their Museums and Their Influence.* Rowman Altamira.

Alexander, V. D., Blank, G., & Hale, S. A. (2018). TripAdvisor Reviews of London Museums: A New Approach to Understanding Visitors. *Museum International*, 154-165.

Anderson, M. L. (1999). Museums of the future: The impact of technology on museum practices. *Daedalus*, 129-162.

Anton M., G. N. (2018). Virtual museums - technologies, opportunities and perspectives. *Romania Journal of Human*, 127-144.

Barbieri, L., Bruno, F., & Muzzupappa, M. (2018). User-centered design of a virtual reality exhibit for archaeological museums. *International Journal on Interactive Design and Manufacturing*.

Barret, J. (2011). *Museums abd the Public Sphere.* West Sussex: Willey-Blackwell.

Batty, M. (2018). *Inventing Future Cities.* Cambridge, MA: MIT Press.

Bautista, S. (2013). *Museums in the Digital Age : Changing Meanings of Place, Community, and Culture.* AltaMira Press.

Bertacchini, E., & Morando, F. (2013). The Future of Museums in the Digital Age: New Models for Access to and Use of Digital Collections. *International Journal of Arts Management*.

Bickham, K. (2016). *British Museum audio guide*. Получено из Kevin Bickham: http://www.kevinbickham.com/work/british-museum-audio-guide/#design-process

Brandta, T., J. B., & Neumann, D. (16 January 2017 г.). Social media analytics and value creation in urban smart tourism. *Elsevier. Information and management*, стр. 703-713.

British museum. (May 2018 г.). *British museum - Private policy*. Получено 21 02 2019 г., из The British museum: https://www.britishmuseum.org/about_this_site/terms_of_use/privacy_policy.aspx

Browne, J. (14 October 2014 г.). *Positioning Visitors with iBeacons*. Получено из BKH Tech: https://www.brooklynmuseum.org/community/blogosphere/2014/10/14/positioning-visitors-with-ibeacons/

Canadian Heritage Information Network. (2012). *Collection Management Software Criteria Checklist 20212*. Получено из Government of Canada: www.canada.ca/en/heritage-information-network/services/collection-management-systems

Capriotti P, L.-D. J. (2018). FACEBOOK AS A DIALOGIC. *El Profesional de la Informacion*.

Carvalho Ana, M. A. (2018). Museum Professionals in a Digital World: Insights from a Case Study in Portugal. *Museum International*.

Cláudia, M. (2012). The museum (audio) guide as an accessibility enhancer. *Accesibilidad en la nueva era de las comunicaciones Profesionales y universidad: un diálogo imprescindible.,*.

Clavir, M. (2002). *Preserving What Is Valued : Museums, Conservation, and First Nations,.* UBC Press,.

Coates, C. (2018). *How Museums are using Virtual Reality.* Получено из https://www.museumnext.com/2019/01/how-museums-are-using-virtual-reality/

Cuau, C. (28 February 2018 г.). *Invisible Insights: learning from TripAdvisor reviews*. Получено из Medium: https://medium.com/mcnx-london/invisible-insights-learning-from-trip-advisor-reviews-b5c825fa4409

Cullinan, N. (25 March 2019 г.). When it comes to museum attendance, it's more than just a numbers game. (T. A. Newspaper, Интервьюер) Получено из https://www.theartnewspaper.com/comment/it-s-not-just-about-the-numbers

Dearnley, L. (2011). Reprogramming The Museum. *Museum and the Web 2011.* Philadelphia.

Doerr, M., & Dionissiadou, I. (2, October 1998 г.). *Data Example of the CIDOC Reference Model. Epitaphios GE34604.* Получено из The CIDOC. Conceptual Reference Model (old version): http://old.cidoc-crm.org/docs/epitafios1.htm

Duin, P. V. (2014). *CLIMATE EFFECTS ON MUSEUM OBJECTS:The Need for Monitoring and Analysis*. Получено из The Getty conservation institute: http://www.getty.edu/conservation/publications_resources/newsletters/29_2/climate_effects.html

E.Hein, G. (1998). *Learning in the Museum.* Routledge.

Ershova, D. (2017). MODERN MUSEUM COMMUNICATION IN THE SPACE OF NETWORK CULTURE. *Society: Philosophy, History, Culture*.

Europeana Pro. (4 March 2015 г.). *Europeana Collections*. Получено из Europeana Pro: https://pro.europeana.eu/services/discovery/collections

Fern Halper, K. K. (2014). *TDWI Big Data Maturity Model Guide.*

Fitzpatrick, L. K. (23 February 2017 г.). *Shared Practices in Museum Open Collections Data*. Получено из Medium: https://medium.com/berkman-klein-center/shared-practices-in-museum-open-collections-data-72e924c4849a

Google Arts&Culture. (March 2018 г.). *MoMA & Machine Learning*. Получено из Experiments with Google: https://experiments.withgoogle.com/moma

Grace, T. (19 Juny 2015 г.). *Improving the Audio Guide: A Look at Our Visitors*. Получено из The Met: https://www.metmuseum.org/blogs/digital-underground/2015/improving-the-audio-guide-a-look-at-our-visitors

Gray, O., Jo Coones, W., & Ruhse, V. (2017). *Museum and Archive in the Move: Changing Cultural Institutions in the Digital Era.* De Gruyter, Inc.

Greater London Authority. (2018). *Smarter London Together.* London: Greater London Authority. Получено из https://www.london.gov.uk/sites/default/files/smarter_london_together_v1.66_-_published.pdf

Hortonworks. (March 2016). *Hortonworks big data maturity model: A Hortonworks White Paper.*

Hsiang-Yi Liu, A. (2008). Using online communities to attract museum visitors. *International Journal of Education through Art*.

ICOM. (November 2011 г.). *Definition of the CIDOC. Conceptual Reference Model.* (C. Nick, M. Doerr, & G. Tony, Ред.) Получено из CIDOC: http://new.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf

Intel Corporation. (2016). *Smart Stadiums lead in Profitability, Fan Experience, and Security.* Intel.

Jarrier Elodie, B.-R. D. (2012). Impact of Mediation Devices on the Museum Visit Experience and on Visitors' Behavioural Intentions. . *International Journal of Arts Management*, 18-29.

Jasper, V. (28 January 2016 г.). *Museums in the city of the future*. Получено из The museum of the future: https://themuseumofthefuture.com/2016/01/28/museums-in-the-city-of-the-future/

Jensen, T. J. (2015). *Museums – Citizens and Sustainable Solutions.* Copenhagen: Danish agency for culture.

John H Falk, L. D. (2012). *Museum Experience Revisited.* Left Coast Press.

Kotler, N. G. (1998). *Museum Marketing and Strategy : Designing Missions, Building Audiences, Generating Revenue and Resources.* San Francisco, CA: Jossey-Bass.

Leissner, D. J. (2016). *Project Overview*. Получено из Climate for Culture: https://www.climateforculture.eu/index.php?inhalt=home

Lin, K. (13 April 2016 г.). *Museum Data And What To Do With It: Carnegie Museums Of Pittsburgh.* Получено из Art Management&Technology Laboratory: https://amt-lab.org/blog/2016/3/museum-data-and-what-to-do-with-it-carnegie-museum-of-art

Madsen, J. J. (2007). HVAC Advice from Museums. *Buildings magazine*.

Malley, J. (11 April 2017 г.). Exclusive: Here's What 3 Big Museums Learn By Tracking Your Phone. *Gizmodo UK*. Получено из https://www.gizmodo.co.uk/2017/04/exclusive-heres-what-museums-learn-by-tracking-your-phone/

Maurer, E. L. (2011). *My Tripadvisor: Mining Social Media for Visitors' Perceptions of Museums vs. Attractions.* Exhibitionist Spring.

Mclean, F. (2012). *Marketing the Museum.* Routledge.

Moriarty, A. (July 23 2018 г.). *AI and Museum Collections*. Получено из Medium: https://medium.com/@adamrmor/ai-and-museum-collections-c74bdb724c07

Moscow authorities. (August 2018 г.). *План развития Москвы 2030/ Умный город будущего.* Получено из Портал правительства москвы: https://www.mos.ru/2030/

OECD/ICOM. (2018). *CULTURE AND LOCAL DEVELOPMENT: MAXIMISING THE IMPACT. Guide for Local Governments, Communities and Museums.*

OECD/ICOM. (2018). *Culture and local development: maximizing effect.* Получено из http://www.oecd.org/cfe/leed/venice-2018-conference-culture/documents/OECD-ICOM-GUIDE-MUSEUMS-AND-CITIES.pdf

Oganesova, J. (2001). PREVENTIVE CONSERVATION OF MUSEUM COLLECTIONS AND ITS ROLE IN THE PRESERVATION OF CULTURAL HERITAGE. *Culturology*, 364-368.

Open Knowledge International . (2007). *The Open Definition 2.1.* Получено из The Open Definition: http://opendefinition.org

Peter Bosch, S. J.-M. (2017). *CITYkeys list of project indicators.* European Commission.

Powerhouse museum. (2017). *API Documentation.* Получено из Museum of Applied Arts and Science: https://api.maas.museum/docs/graphql

Ruth Rentschler, A.-M. H. (2009). *Museum Marketing.* Business & Economics.

Sandell R., J. R. (2007). *Museum management and marketing.* Routledge.

Silverman, L. H. (2010). *The Social Work Of Museums.* London: Routledge.

Simnick, A. (January 2017 г.). The power of applied data for museums.

Slingerland, G. (27 November 2017 г.). *"PARK MAKERS": A DIGITAL PLATFORM THAT SUPPORTS CITIZENS COMMUNITIES USING OPEN DATA*. Получено 01 05 2019 г., из Medium: https://medium.com/open4citizens/park-makers-a-digital-platform-that-supports-citizens-communities-using-open-data-3eedd00fcaa1

Sonoda, N. (2012). *Managing and Analyzing Museum Environmental Data.* Tokio.

Strelka institute. (18 02 2016 г.). МУЗЕИ XXI ВЕКА: МЕЖДУ ВИРТУАЛЬНОЙ РЕАЛЬНОСТЬЮ И ТОРГОВЫМ ЦЕНТРОМ. *Strelka Mag*. Получено из https://strelkamag.com/ru/article/musei-discussion

Styliani S, F. L. (2010). Virtual museums. a survey and some issues for consideration. *Journal Cultural Heritage*, 520-528.

Sue Runyard, Y. F. (1999). *The Marketing and Public Relations Handbook for Museums, Galleries and Heritage Attractions.*

Suzić, B. K. (2016). Adoption of Social Media for Public Relations by Museums. *Central European Business Review*.

The Pushkin museum of Fine Arts & JSKT data group. (2017). *Музей.Пространство внимания*. Получено из Heatmap.Arts-museum: http://heatmap.arts-museum.ru

The Pushkin State Museum of Fine Arts. (б.д.). *Коллекция*. Получено из Официальный сайт Пушкинского музея: https://www.pushkinmuseum.art

W3C. (February 25 2014 г.). *RDF 1.1 Concepts and Abstract Syntax.* Получено из W3C Working Group Note: https://www.w3.org/TR/rdf11-concepts/

WALKER, K., & TALLON, L. (2009). *Digital technologies and the museum experience: Handheld guides and other media.*

Wazlawick R. S. Rosatelli, M. C. (2001). Providing More Interactivity to Virtual Museums: A Proposal for a VR Authoring Tool. *Presence: Teleoperators & Virtual Environments*, 647-656.

Ylva French, S. R. (2011). *Marketing and Public Relations for Museums, Galleries, Cultural and Heritage Attractions.* Routledge.

Zafiropoulos, K., Vrana, V., & Antoniadis, K. (2015). Use of twitter and Facebook by top European museums. *Journal of Tourism, Heritage & Services Marketing, 1*, 16-24.

Zanibellato, F., Rosin, U., & Casarin, F. (2017). How the Attributes of a Museum Experience Influence Electronic Word-of-Mouth Valence: An Analysis of Online Museum Reviews. *International Journal of Arts Management.* , 76-90.

КАМИС. (2018). *Функциональные характеристики КАМИС*. Получено из КАМИС.Музейные системы: https://docviewer.yandex.ru/view/642160633/?*=8P0b%2Bg3X6t%2FF8S7f%2Fuiggbsl7Sp7InVybCI6InlhLWJyb3dzZXI6Ly80RFQxdVhFUFJySlJYbFVGb2V3cnVMU3NObG5KeVdXcWZqUWxyTkNkalNRRnU4SE5mTDlpbGNmWTFxTVB5TGxNa1p6dFFUMkxuYnZHWmZORXRNSTl4dUtqSDc5dG1RZmkxOC0xZjYxQTUtalQwUFU

П. О. Васильева, Д. В. (2018). *Музей в цифровую эпоху: перезагрузка.* Москва: Издательские решения.