

UCL QATAR

كلية لندن الجامعية قطر



UCL

MUSEUM BIG DATA

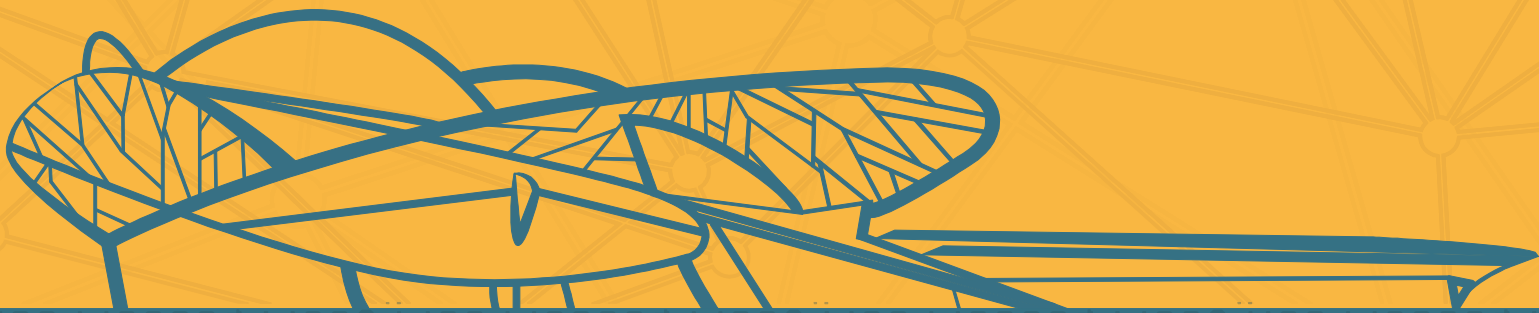
Practices and Prospects



MUSEUMS AND BIG DATA

1ST INTERNATIONAL CONFERENCE

30th April - 2nd May 2019



QNL Auditorium
April 30th

2:30 - 6:30 PM

UCL Qatar, room 1D02
May 1st – 2nd

9:30 AM - 3:30 PM



MUSEUM BIG DATA



Head of Conference: Dr Georgios Papaioannou, UCL Qatar

Head of Organising Committee: Aurelie Jouvenel, researcher, UCL Qatar

Scientific Committee: Dr Georgios Papaioannou, UCL Qatar, Dr Milena Dobрева, UCL Qatar, Aurelie Jouvenel, UCL Qatar

The Conference is part of the Museum Big Data Project, funded by Qatar Foundation / UCL Qatar.

Special thanks for their contributions to Qatar National Library, the Sheikh Faisal Bin Qassim Al Thani Museum, the Media Majlis at Northwestern University in Qatar , The National Museum of Qatar and the UCL Qatar staf.

Conference Programme

April, 30th

○ Visit Sheikh Faysal Museum

○ Qatar National Library, Auditorium

**Collecting, storing and data access:
Service development and its challenges**

14:30 – 15:00

Introduction

15:00 – 15:20

Mahendra Mahey (The British Library) *"Examining practices and prospects of using and sharing big data in 'Galleries, Libraries, Archives and Museum (GLAM) digital Labs'"*

15:20 – 15:40

Sarah Stewart (The British Library / University of Oxford)
"Towards a Networked Digital Cultural Heritage: Data Services and Persistent Identifiers at the British Library"

15:40 – 16:00

Armin Straube (UCL-Qatar) *"The digital convergence of information professionals: Common challenges for museum curators, librarians and archivists in the digital realm"*

16:00 – 16:30

Discussion

16:30 – 17:00

○ **Coffee break**

17:00 – 17:20

Milena Popova (Europeana) *"The ripple effect of big cultural data: applications in education, research and more"*

2

17:20 – 17:40

Susan Reilly (Qatar National Library)

"Memories written in sand and water: building the digital memory of Qatar"

17:40 – 18:00

Gail Feigenbaum (The Getty Research Institute)

"Search and Research: Making Art Data History"

18:00 – 18:30

Discussion

Dinner at QNL

May, 1st

UCL Qatar

Talking of methodology: Data visualisation and Textual analysis

9:30 – 9:45

Welcoming by Sam Evans, director of UCL-Qatar

9:45 – 10:05

Theodosia Prodromou (University of New England)

"Data Visualization and Statistical Literacy for Open and Big Data"

10:05 – 10:25

Joan Beaudouin (Wayne State University)

"Hidden in plain sight: Visualizing collection holdings of The Henry Ford Museum of American Innovation" followed by "Online art museum collections in the United States: Assessing their public face"

10:25 – 10:55

Discussion

10:55 – 11:15

Coffee Break

11:15 – 11:35

Vladimir Alexiev (Ontotext)

"Museum Linked Open Data: Ontologies, Datasets, Projects"

Conference Programme

11:35 – 11:55

Lonneke Van der Plas (University of Malta) *"Natural Language Processing for Museum Data"*

11:55 – 12:15

Georgios Papaioannou and Aurelie Jouvenel (UCL-Qatar) *"Sentiment analysis for Museums: SensioM"*

12:15 – 12:45

Discussion

12:45 – 13:45

Lunch Break

Audience research

13:45 – 14:05

Chiara Zuanni (University of Graz)
"The Big Data Museum: exploring digital flows around heritage objects, institutions, and audiences"

14:05 – 14:25

Ashley Burgoyne (University of Amsterdam) *Hooking the Public with Hooked on Music*

14:25 – 14:45

Georgios Papaioannou (UCL-Qatar), Aurelie Jouvenel (UCL-Qatar), Kenza Ziar (Qatar Museums), Mariel Balagtas Cunanan (Qatar Museums) and Anne Sophie Marie Brault (Qatar Museums), *"Museum Big Data and Audience Research in museums in Qatar: a collaborative project between UCL Qatar and Qatar Museums"*

14:45 – 15:15

Discussion

**Visit Media Majlis at
Northwestern University
in Qatar**

Dinner at Damasca

May, 2nd◦ **UCL Qatar****Strategies, decision making
and legal issues****9:30 – 9:50****Kenza Ziar (Qatar Museums), Mariel Balagtas Cunanan (Qatar Museums)
and Anne Sophie Marie Brault (Qatar Museums) *title to be defined*****9:50 – 10:10****Géraldine Salord (Paris Bar – METALaw Firm) *“Legal issue of Big Data:
using law as a tool to build a Big Data strategy in the cultural field”*****10:10 – 10:40****Discussion****10:40 – 11:10**◦ **Coffee break****11:10 – 11:30****David Haskiya (Swedish National Heritage Board)
*“Challenges and opportunities for Swedish museums in digitising
their collections”*****11:30 – 11:50****Angie Judge (Dexibit)
*“Predicting the future for museums using big data and AI”*****11:50 – 12:20****Discussion****12:20 – 13:00****Conclusion****13:00 – 14:30**◦ **Lunch****Visit of the NMoQ
Dinner at Parisa**

Museums nowadays produce data on and by their objects (cataloguing, curating, conserving, digitising), their operational activities (daily functions, communications, management, research, marketing, other operations), and their visitors (museum/heritage institution visits, other activities, educational activities, visitors' involvement/input and feedback, and social media). This is a huge amount of information in an era of data explosion. This is (or can be) Big Data. Big Data have come to describe this new generation of data characterised by extended volume, velocity, variety and visitors. Big Data, together with analytics and data mining, are a valuable tool and a driver for improving products, understanding behaviours, shaping strategies and policies, developing new products, ideas and services. Within this setting, the cultural sector, including museums and heritage institutions, faces a time of exploration and updating. Museums learn from, and can teach, other institutions, such as libraries, archives and other heritage, information and memory institutions. All those institutions are multi-character and have activities and operations ranging from information and conservation/preservation to dissemination, education and fun.

In this conference, we explore the meaning of Big Data, their links to museums and heritage institutions, and certain projects related to museums and Big Data, aiming to describe the presence of Big Data in the world of museums and heritage institutions.

The Conference will:

- identify common grounds and differences in Big Data practices among different institutions, including museums, heritage/cultural institutions, libraries, archives and other information/memory institutions,
- put museum and heritage institutions professionals and data scientists together and to engage in a debate on how big data analytics and analysis can contribute to decisions making in museums and heritage institutions,
- present and exchange ideas and observations on methodology, practices, ethics and other related issues with researchers and professionals from all over the world,
- share results from different projects around the world,
- address the impact of Big Data use on museum and heritage institutions' evaluation processes.

Collecting, storing and data access: Service development and its challenges

Mahendra Mahey has been leading and managing British Library Labs ('BL Labs'), from its inception in 2013. Based in the Digital Scholarship department at the British Library, it was previously an Andrew W. Mellon foundation funded activity. BL Labs is now transitioning into a British Library supported sustainable service to provide access to and expertise on inspiring the use of and experimentation with its vast and unique digital collections and data in new, exciting and innovative ways.

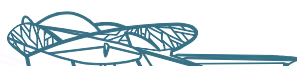
'BL Labs' helps researchers, artists, entrepreneurs, educators and innovators to work on, experiment, incubate and develop their ideas of working with the BL's digital content through competitions, awards, projects, exhibitions and other engagement activities. It does this by providing services and infrastructure to enable, facilitate and give access to its data openly and onsite for research, inspiration and enjoyment for everyone in the world. BL Labs provides a platform to promote and celebrate success in projects and communicate how lessons are learned from how challenges are addressed and overcome.

Mahendra has been sharing his knowledge and experiences through an international community with his colleagues who are helping to build it with Galleries, Libraries, Museums

and Archives (GLAMs) that either had, are planning or already have digital experimental 'Labs'. By generously sharing and 'paying forward' their expertise, knowledge and experiences the group hope to ensure that organisations don't have to 're-invent the wheel', they can learn from each other and enable collaboration across borders through their digital collections, data, services, infrastructure and practice. This group aims that this work will result in building better digital 'Labs' for their organisations and their users and help to further open up their data and services for everyone, from any background in the world to use on projects. It is hoped that this will create a self-sustaining continuous and virtuous cycle to constantly inspire others to use and reuse digital collections, providing a return on the considerable investment organisations have made in preserving and providing access to their digitised and born digital content.

For the last 20 years, Mahendra has built a strong reputation of working with digital technology and data as a manager, educator, adviser and community builder in Further and Higher Education, cultural heritage, businesses and other sectors for researchers, artists, educators, curators, entrepreneurs, software developers, librarians and other professionals both in the UK and internationally.

Title: *Examining practices and prospects of using and sharing big data in 'Galleries, Libraries, Archives and Museum (GLAM) digital Labs'*



Experiences and lessons learned from the British Library and organisations around the world engaging with researchers, artists, educators and entrepreneurs who have used digitised cultural heritage collections and big data.

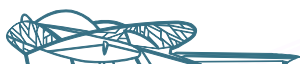
Abstract: The British Library is one of the largest national libraries in the world and is creating and storing millions of digital items every year such as digitised books, newspapers, maps, sheet music, manuscripts, audio / TV recordings as well as born digital archived websites, personal digital archives, electronic books, radio, performances, and artworks. This incredible range of digital material is having a profound effect on the way our libraries are supporting those who want to use digital content and methods in their work. What new facts will scholars discover when they analyse thousands of digitised books computationally using data-mining techniques? What are the challenges and solutions for libraries to build systems and services that provide seamless access to its digital material from a radio recording to newspaper story? What are the practical experiences of working on digital crowdsourcing projects, and how is machine learning helping libraries to unlock new information hidden in its digital archives? Can we use digital technologies to visualise and shine light on a library's holdings, and unearth unusual and surprising findings artistically? Mahendra will give a brief overview of digital collections and data being made available through British Library Labs ('BL Labs') and examine how they have been re-used by making connections and collaborating with digital researchers, artists, entrepreneurs, educators, curators and librarians around the world through a range of innovative

projects, research questions and engagement activities. He will highlight the myths and assumptions many make about GLAMs and address the significant issues and challenges they face when working with digital collections and data (e.g. legal, technical, human etc.). He will reflect on lessons he has learned over nearly two decades of working in Further and Higher Education, cultural heritage and other sectors, suggesting the types of digital research that could bring significant benefit and impact to the way GLAMs in particular may work into the future.

Mahendra will describe an exciting international support community that he is starting to build with colleagues around the world. This community is bringing national, state, university and public Galleries, Libraries, Archives and Museums together that either had, are planning or already have experimental digital 'GLAM Labs' which encourage their users to re-use their digital collections and big data. The group are already providing support to each other, sharing expertise, knowledge and experience and are pooling resources together in order to build better innovative digital 'GLAM Labs' that bring value to their organisations and users well into the future. They also intend to publish a practical open access book in 2020 about 'Building better GLAM Labs' which will include a description of the landscape of digital 'GLAM Labs' worldwide and provide advice and guidance for GLAMs which are in process of setting up or currently running digital 'GLAM Labs'.

.....

Sarah A. Stewart is a Data Services Specialist at the British Library and a DPhil student at the Oxford Internet



Institute, where she is investigating the impact of open data on research practices within natural history collections. She has a strong interest in digital scholarship, digital cultural heritage and open research.

Title: *Towards a Networked Digital Cultural Heritage: Data Services and Persistent Identifiers at the British Library*

Abstract: The British Library is the national library for the United Kingdom, holding over 170 million items and counting. While only 3 percent of the collections have been digitised to date, the library still holds a substantial digital collection which now includes research datasets and software. The British Library is currently developing a shared research repository service which will enable search and discovery of the research produced by its staff. It has also seen the development of a data services team, which will enable datasets and software developed from research at the British Library become FAIR (Findable, Accessible, Interoperable and Reusable). This talk will also present the application of Digital Object Identifiers (DOIs) and other persistent identifiers for a networked digital cultural heritage.

.....

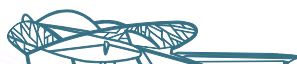
Armin Straube is Teaching Fellow in Library and Information Studies at UCL Qatar. Before joining UCL he worked as Head of Data Curation at Qatar National Library and as manager of nestor (the German Network for Digital Preservation) at German National Library.

He was involved in a number of

projects in the wider cultural domain, contributing to digitization efforts, the standardisation and enrichment of descriptive metadata and the development of data models and metadata exchange formats. His experience includes consultancy and teaching assignments in the areas of archives and records management, digital archives and research data. Armin holds master degrees in history (University Hagen, Germany), in geography (University Halle, Germany) and in archives and records management (University College Dublin, Ireland). His research interests include data curation and digital preservation and he is looking into the question how the changing professional practice can be reflected in the teaching of information science.

Title: *The digital convergence of information professionals: Common challenges for museum curators, librarians and archivists in the digital realm.*

Abstract: Information professionals working in museums, libraries, archives and research institutions face similar challenges managing the digital holdings in their care. The growing number, diversity and complexity of the information objects and the fast pace of technological change require the constant (re-) development of infrastructures that enable access and ensure digital preservation. Access itself has become more complex in the digital realm. The distinction between description and object, between metadata and data becomes blurred as digitized and born digital material is used in new and unprecedented ways. Machine learning and big data



Applications help to uncover new patterns and insights from collections. The aim to linking items semantically beyond the confines of collections or institutions and the drive to increase visibility through participation in aggregation projects like Europeana or the Digital Public Library of America add further layers of complexities. Museums, libraries and archives contribute to these initiatives and bring their unique content, approaches and insights. These joint efforts however also illustrate the increasing similarities in challenges faced and solutions implemented across all memory institutions.

Regardless of the institution type, information professionals need to be able to:

- make appraisal decisions,
- record provenance and context,
- capture or generate descriptive and technical metadata in various formats,
- manage very diverse digital material and content,
- understand digital preservation needs and give input into system development and
- prepare data for a multitude of (new) use cases.

The presentation will detail how information professionals across all memory institutions can meet these challenges, learn from each other and jointly make their collections and knowledge available to all.

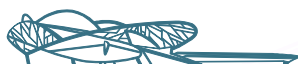
Milena Popova work as a Business Development Manager at Europeana Foundation. She leads the Re-use services team which aims to increase

the use of digital cultural data in education, research and by the creative industries. Milena is responsible for the development of distribution partnerships with diverse networks (from innovative labs and start-up hubs to educational associations, publishers and research infrastructures). She also oversees the promotion and uptake of Europeana incubation services (challenges, match funding) and the growth of the market specific communities. Prior to this position, Milena worked for German and Bulgarian software companies in marketing and business development roles.

Title: *The ripple effect of big cultural data: applications in education, research and more*

Abstract: Large, complex and fast growing volumes of data are the signature mark of modern life and can be found in every industry, from manufacturing to finance, healthcare, social media and culture. Collecting, storing and managing such big data comes with its own great challenges; however, it also brings tremendous opportunities to extract useful insights and create new forms of value.

This session will introduce Europeana, Europe's digital cultural heritage platform, as an inspiring showcase of big cultural data. The platform currently provides access to nearly 60 million digital cultural records from over 3,500 libraries, museums, archives and audio-visual collections across Europe. Europeana Collections feature a great thematic, media and language variety: one can browse content on topics from arts, history, geography and music to newspapers, films, and more; in all media formats (image/text/



sound/ video/3D) and in over 35 languages.

The presentation will show how building up and operating with these huge, rich and diverse collections has helped drive digital transformation and innovation not only in the cultural heritage sector but also in education, research and even daily life.

.....

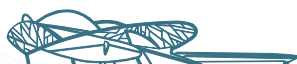
Joan E. Beaudoin is an Associate Professor in the School of Library and Information Science. Prior to coming to Wayne State University she was a Laura Bush 21st Century Librarian Fellow at the School of Information Science and Technology at Drexel University. While at Drexel University she undertook several research projects that examined users' description of images. Her dissertation "An investigation of image users across professions: A model of image needs, retrieval and use," expanded upon these earlier projects and examined the image behaviors of archaeologists, architects, art historians and artists. Her research interests developed from the practical experiences she had teaching art history, performing archaeological field work, and working in the field of visual resources. In addition to a Doctor of Philosophy in Information Studies at Drexel University, she holds a Master of Science in Library and Information Science degree in the Management of Digital Information from Drexel University, a Master of Arts in art history from Temple University, and a Bachelor of Fine Arts in art history from Massachusetts College of Art.

Title: *Hidden in plain sight: Visualizing collection holdings of The Henry Ford Museum of American Innovation*

Abstract: Collection data describing the contents of museums plays a pivotal role in the reception of the objects held by an institution. This is particularly the case in the presentation of online museum collections. Using collection data from The Henry Ford Museum of American Innovation this presentation looks at how information visualization techniques can reveal hidden aspects about collection holdings. Various visualizations of the museum's data are presented to explore what can be learned, and how this higher level of access differs from what is currently available for online museum collections. As visualizations can foster innovative interactions with cultural materials, multiple audiences would benefit from this additional layer of discovery. Who these audiences are, and the potential advantages of providing data visualizations of collections are investigated.

Title: *Online art museum collections in the United States: Assessing their public face*

Abstract: As all art reflects the time and place in which it was created, the underlying mission of art museums is to engage, instruct, inspire, and enrich lives. Art inevitably forces each of us to look at our own lives and reflect upon our role in our shared human history. While museums contain rich stores of objects, the wealth of their resources is largely inaccessible to the public due to a multitude of barriers (e.g., geographical, financial, and time). Online collections can extend the multiple important functions of museums to many individuals beyond the confines of museum walls. Thus, this presentation examines the availability of, and ease of access to, art museum collections presented through web-based collection systems. Focusing on art museum collections in the United States,



this project analyzes the data behind the representation of museum objects in the online setting to codify and report on shared practices. Questions concerning fundamental features of online collection access, such as what percentage of the collections are available online, and semantic commonalities found across the museums are explored, and allow us to reflect upon how far museums have extended their reach.

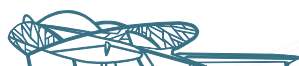
.....

Gail Feigenbaum is associate director of the Getty Research Institute where she oversees publications and research projects. A scholar of early modern European art, especially Italian and French painting, she published extensively on the Carracci family of painters, on Caravaggio, and Georges de la Tour. She holds a doctorate from Princeton and began her career at the National Gallery of Art in the education and curatorial departments, and at the Center for Advanced Study in the Visual Arts. Her recent books are *Display of Art in the Roman Palace*, *Provenance: An Alternate History of Art*, and *Sacred Possessions: Collecting Italian Art*. Her fellowships and awards include the Rome Prize at the American Academy, CASVA, NEH, and the Tessin Medal, National Museum of Sweden. Prior to joining the Getty she was curator of Painting at the New Orleans Museum of Art where she organized *Degas and New Orleans: A French Impressionist in America* and *Jefferson's America and Napoleon's France: An Exhibition for the Louisiana Purchase*. Among her exhibitions are *Ludovico Carracci*, Kimbell Art Museum and *Pinacoteca Nazionale, Bologna*, and *The Drawings of Annibale Carracci*, National Gallery of Art. Her

current research project is on *America and the International Art Market, 1880-1930*.

Title: *Search and Research: Making Art Data History*

Abstract: Arts institutions, like the Getty, began to undertake big--or comparatively big--data projects several decades ago. A handful of these groundbreaking projects still persist in some form, as will be discussed here, and a few have even turned forty. The digital environment in art history and museums has evolved. It certainly has grown in emphasis and scale. Periodically the field pauses to take stock, and many of us have experienced recurrent institutional reorganizations of our digital programs. We have experienced a succession of "next big things," ambitious and much vaunted projects, systems, concepts or technologies, including some that have faded into obscurity. There has been an incredibly intensive commitment of human and financial resources to digital initiatives that have revolutionized how we do research on art and make it accessible to audiences. I argue that the phenomenon of digital art information by now has acquired a history. In this aggregation of projects unfolding over time, patterns of behavior, intention, and reception have become evident, and are susceptible to analysis. The digital environment is insistently the now and the future; by contrast, history affords a critical point of view, offers a perspective on the present situation. Despite constant innovation, and the rapid churn of staff and technology, we are continuously building on yesterday's concepts and yesteryear's software. Toward a history, a partial sketch of the Getty's pioneering endeavors in digital art information, and early articulation of principles for



an integrated architecture of art information of the future is offered here. Of the many projects, a few recent ones are introduced, while greater focus is given to two long duree survivors: the Getty Vocabularies and the Getty Provenance Index.

Talking of methodology: Data visualisation and Textual analysis

Dr Theodosia Prodromou (Ph.D., M.Sc., M.A., BSc (Hons), Dip.Teach., Dip.Art., G.C.T.E.) is a mathematician, statistician and mathematics educator. She is currently a Senior lecturer at the University of New England in Australia. She has a Ph.D. in Statistics education from Warwick university in United Kingdom, an M.Sc. in Statistics (Warwick university, United Kingdom), B.Sc. in Mathematics and Statistics (Aristotle University, Thessaloniki, Greece). She worked as a statistician and secondary Mathematics teachers in different countries of Europe, and Australia. She has experience in teaching statistics to undergraduate students or professionals who do not have a strong mathematical background. She has also experience in teaching mathematics education to pre-service teachers and in-service teachers within secondary and post-graduate programs. She is working on numerous research projects worldwide related to technology integration into schools; STEM education; Secondary teachers' professional development; statistics education, statistical literacy, Big Data and Big Data Analytics in Educational settings.

She has engaged extensively in research

and has established a respected record through numerous publications in scholarly journals, book chapters and books.

Title: *Data Visualization and Statistical Literacy for Open and Big Data*

Abstract: The recent developments in data technology allows as to have scalable and flexible data capture, storage, processing and analysis.

The emerging field of new types of “big data” and “big data analytics” is a growing field of research that informs the process of using data to improve learning and teaching at the institutional level and transform the way applied social science is done.

After reviewing the notion and characteristics of Big Data and big data analytics, I will discuss how big data analytics could be used in educational settings, in libraries, arts and humanities. I will discuss how big data analytics may be exploited to improve learning outcomes of undergraduates in higher institutions towards attaining sustainable education programmes at Universities.

I will also present case studies from the chapters of my two authored/edited books entitled: “Data Visualization and Statistical Literacy for Open and Big Data” (2017, Hershey: IGI Global) and “Big Data in Education: Pedagogy and Research (in press/ New York: Springer. These examples of Big data and Big Data analytics in education, pedagogy and research at Humanities, Arts, Social Sciences and Education will illustrate prerequisites and potentials of improving research outcomes,



new ways of visualizing and presenting complicated information.

.....

Vladimir Alexiev has a PhD in computing science from University of Alberta, MS in computer science from Technical University of Sofia, PMP certification, Project Risk and Quality Management diploma. He has 30 years of IT experience, of which 20 years of IT PM experience and 10 years semantic technologies experience.

He is one of the founders of Sirma Group Holding, the largest Bulgarian IT group, and the parent company of Sirma.AI (Ontotext). He joined Ontotext in 2010. Vladimir is a Masters-level lecturer in IT PM at New Bulgarian University. His experience includes ontology engineering, metadata standards, vocabularies and thesauri, RDF, RDFS, OWL2, RDF Shapes, SHACL, SKOS, SPARQL, LOD, mapping, R2RML, ETL, semantic web applications, project management, business analysis and requirements specifications. He has worked in various business domains, e.g. Customs and Excise, Personal Finance workflows, Legal Procedures and Statistics, Cultural Heritage and Digital Humanities, Company data (Trade Registers, Mergers and Acquisitions, Startups), commodity trading info (incl. Oil & Gas and Energy), Science data (researchers, organizations, projects, programs, publications, impacts), Health Records (FHIR), etc.

In Ontotext he leads the Data and Ontology Management practice. He has lead a wide range of projects

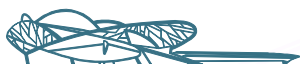
related to data integration, linked open data, cultural heritage, LOD management, taxonomies/thesauri, etc. He served on 5 Europeana task forces (including 2 on semantic enrichment), the Members Council and the Data Quality Committee. He is on the DBpedia Ontology and Data Quality committee. He has contributed to the ontology definition of the

ISO 25964 standard for thesauri management; on publishing the important Getty Vocabularies thesauri as LOD, Wikidata property definitions and authority dataset coreferencing. He has consulting experience with complex ontologies such as CIDOC CRM, FIBO, Schema.org; optimization of inference rules; faithful semantic representation of humanities and CH data; CH aggregators.

CV and Publications: <http://github.com/Vladimir-Alexiev/my/>

Title: *Museum Linked Open Data: Ontologies, Datasets, Projects*

Abstract: The Galleries, Libraries, Archives and Museums (GLAM) sector deals with complex and varied data. Integrating that data, especially across institutions, has always been a challenge. Semantic data integration is the best approach to deal with such challenges. Linked Open Data (LOD) enable large-scale Digital Humanities (DH) research, collaboration and aggregation, allowing DH researchers to make connections between (and make sense of) the multitude of digitized Cultural Heritage (CH) available on the web. An upsurge of interest



in semtech and LOD has swept the CH and DH communities. An active Linked Open Data for Libraries, Archives and Museums (LODLAM) community exists, CH data is published as LOD, and international collaborations have emerged. The value of LOD is especially high in the GLAM sector, since culture by its very nature is cross-border and interlinked. We present interesting LODLAM projects, datasets, and ontologies, as well as Ontotext's experience in this domain. An extended paper on these topics is also available. It has 77 pages, 67 figures, detailed info about CH content and XML standards, Wikidata and global authority control.

.....

Lonneke van der Plas is senior lecturer at the Institute of Linguistics and Language Technology of the University of Malta since October 2014, where she is the chair of the European Masters program Erasmus Mundus in Language and Communication Technologies. Before that, she was junior professor at the Institute for Natural Language Processing (IMS), University of Stuttgart, did her a post-doc at the University of Geneva working in the field of cross-lingual transfer of semantic role labelling as part of the CLASSiC project, and earned her PhD from the University of Groningen, where she worked on automatic lexical acquisition from corpora. She has over 50 publications in the field of natural language processing, more in particular on the following topics: computational linguistics, cross-lingual natural language processing, distributional semantics, text mining, terminology extraction, question answering, semantic role labelling, tools for low-resource languages, diachronic computational linguistics, word formation

(compounding), computational creativity.

Title: *Natural Language Processing for Museum Data*

Abstract: In this talk, I will give an overview of the field of natural language processing (NLP) focusing on its relevance for big data related to museums. I will show where we currently are in NLP, what our state-of-the-art tools can do, while also showing some limitations. I will talk about social media text, and how one can harvest it to analyse emotions and extract opinions people express about a specific topic. I will show how social media text differs from standard newspaper text and what challenges this gives rise to in terms of the automatic processing of these texts. Lastly, I will talk about an understudied topic in NLP, which is computational creativity. I believe that museum content in all its richness and diversity is particularly suitable for inspiring people, and this potential can be unlocked by creating tools that find patterns and make connection in museum content. I will briefly show some results from ongoing work I am pursuing with respect to computational creativity in historical data.

.....

Georgios Papaioannou is an Associate Professor for the MA in Museum and Gallery Practices at UCL-Qatar. He has lectured, excavated, led tours and conducted museum / cultural heritage work in the UK, Greece, Cyprus, Australia, Spain, Egypt, Syria, Oman, Turkey, Yemen and Saudi Arabia. He has set-up exhibitions in Jordan (the Museum at the Lowest Place on Earth, Dead Sea), Oman, Sweden and at various museums in Greece. He has led and



coordinated multi-partner cross border research and innovation projects, and he received project coordination international awards. Dr Papaioannou is General-Secretary of the Hellenic Society for Near Eastern Studies, Director of the Museology Lab in Corfu (Greece) and a member of ICOM. He leads a project financed by Qatar Foundation on Big Data and Museums.

Dr Papaioannou has studied archaeology, classics, cultural heritage and IT in Greece (Ioannina, BA), in Britain (MA at UCL, PhD at King's) and in Spain (Tenerife). His research interests lie in museology, archaeology of the Eastern Mediterranean and the Arab World, education (including e-learning), tourism (including city tourism), cultural studies and IT applications, including virtual reality, augmented reality and mobile applications. He is currently writing a book on museology as a discipline of information sciences.

Aurelie Jovenel is a Research Assistant in Museum and Gallery Practices at UCL-Qatar. She has collaborated with Dr Alexandra Bounia on a research programme on collecting from contemporary traumas and with Dr Georgios Papaioannou on Big Data and Museums producing articles and organising conferences. Those two projects combine some of her interests: identity, cultural trauma and national construction as well as mathematical tools in cultural area and knowledge transmission. She has studied art history and archaeology in France (Ecole du Louvre, MA at Lyon 2) and she is finishing her PhD *Territories and identities in Southern Levant at the beginning of the Iron Age* at the EPHE, Paris, under the supervision of Dr. Robert Hawley.

Title: "Sentiment analysis for Museums: SENSIOm"

Abstract: The SENSIOm is a data dashboard on museum visitors-related datasets for sentiment analysis. It has been one of the outcomes of the Museum Big Data Research Project in Qatar, initiated in 2017 at University College London in Qatar led by the authors. It is a dynamic multi-functional data dashboard on Museum Big Data to serve research needs on Negative/Neutral/Positive sentiment analysis and analysis on museum visitors' views on different museum-related topics, such as price of tickets, queue, other assets and issues. Sentiment analysis is benchmarked against other museums and/or cultural institutions and hotspots. As part of our Museum Big Data Project, we aim to create a digital path for museums to process unbiased feedback in digital format from museum real and/or online visitors, and to analyse it with regards to visitors-generated multi-lingual data, comments and sentiments, along with a dynamic data dashboard representing market benchmarks from direct and indirect competition. In this paper we present the processes we follow to train the SENSIOm on museum visitors-related datasets for sentiment analysis.

Audience research

Dr Chiara Zuanni is a tenure-track assistant professor in Digital Humanities, museology-focused in the Centre for Information Modelling – Austrian Centre for Digital Humanities at the University of Graz. Her research focuses on the construction and mediation of knowledge in museums,



the impact of digital media on the heritage sector and its audiences, and digital data practices in museums. She is a member of the Institute in Ancient Itineraries (2018–2019, funded by the Getty Foundation, led by King's College London and Umeå University) and responsible for the University of Graz of the Erasmus+ project 'DigiCulture' (2018–2021, coordinated by the University of Timisoara).

Title: *The Big Data Museum: exploring digital flows around heritage objects, institutions, and audiences*

Abstract: This paper will discuss the production and flows of data in museums, focusing on the implication of this data for museum practices, for digital engagement, and for digital preservation. On the one hand, the digitisation of the collections and the digitalization of museum work produce a huge, unstructured, and not yet fully appreciated amount of data and data practices. On the other hand, the mediation and consumption of heritage objects in the digital sphere contributes to new forms of participation, in which data are generated by the museum, by its audiences, and in their interactions.

This paper aims to examine the range of data, their flows, and the practices surrounding them in the museum. It will begin by observing the datafication of museum content and processes, and its focus will then be on user-generated data to highlight how visitors' contributions during the museum visit or online reaction to the museum content or activities generate new flows of data that, in turn, reveal audiences' expectations, perceptions, and attitudes towards the museum itself. By drawing on

case studies analysing viral museum posts, the paper will discuss the potential of data science practices for researching museum audiences and the methodological and ethical implications of this approach.

At the same time, the paper will highlight the value of preserving all these contributions, their content, their metadata, and their stories, since they represent particular time-points in the history of an object and are valuable source for future museum and cultural historians. In order to do so, the paper will therefore discuss the need of developing data cultures within museums and the heritage sector, as well as an increased understanding of digital preservation and archival practices

.....

John Ashley Burgoyne is the Lecturer in Computational Musicology at the University of Amsterdam and part of the Music Cognition Group at the Institute for Logic, Language, and Computation. Dr Burgoyne teaches in both musicology and artificial intelligence and is especially interested in musicometrics: developing behavioural and audio models that are conceptually sound, reliable, and musicologically interpretable as music enters the digital humanities era. He was the leader of the Hooked on Music project, an online citizen science experiment to explore long-term musical memory that attracted more than 170,000 participants across more than 200 countries.

Title: *Hooking the Public with Hooked on Music*

Abstract: Music is fundamental to many people's identities and remains a strong trigger for memories from youth to senescence – but not all music. What is are



musical differences between those sounds that lodge themselves in our memories after only a few hearings and those we quickly forget? The Music Cognition Group at the University of Amsterdam, in partnership with Utrecht University, the Netherlands Institute for Sound and Vision, and the Manchester Museum of Science and Industry, undertook Hooked on Music, a citizen-science experiment to uncover the musical characteristics that are most responsible for long-term musical memory. The project focused on the music successful popular music in the United Kingdom from the 1940s through to the present – and attracted more than 170 000 participants to engage with the museum in some way. This talk will discuss some practical aspects of what did (and didn't) work in driving public engagement, describe the listening strategies that participants used to recognise popular music, and introduce an example of how the results of this project are currently being used to work to enhance an archival collection of 14 000 78-rpm recordings of Dutch popular music from the mid-20th century.

Strategies, decision making and legal issues

Georgios Papaioannou (UCL-Qatar), Aurelie Jouvenel (UCL-Qatar), Kenza Ziar (Qatar Museums), Mariel Balagtas Cunanan (Qatar Museums) and Anne Sophie Marie Brault (Qatar Museums)

Title: *Museum Big Data and Audience Research in museums in Qatar: a collaborative project between UCL Qatar and Qatar Museums*

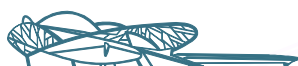
Abstract: This paper presents a new project on museum Big Data and data mining using data related to museums in Qatar as a case

study. In museums in Qatar, as elsewhere in the world, there is an emerging need to detect new and discover hidden and useful information, patterns, clusters and relationships among large sums of museum-related data. Addressing this need requires ethical considerations and processes, a thorough understanding of contexts in the real and the digital world, and cross-disciplinary Big Data methods, techniques and testing, all of which fall within this new project's objectives and discussion points demonstrated here for the first time.

The aims of this Research Project, which is a collaboration between University College London in Qatar and Qatar Museums, are to contribute to the development of Big Data and Data Mining methods and techniques on museum datasets, to add to a policy document on Big Data and the museums in Qatar, to establish a research team on Museum / Cultural Heritage Big Data in Qatar, to explore links / collaborations to information seeking research schemes related to Social Media cultural dataset-producing processes in Qatar. The project awaits ethical approvals from London and Qatar to be officially launched.

.....

Géraldine Salord obtained a Ph.D in Law from Paris II-Panthéon-Assas University (2007), before being admitted to the Paris Bar (2010). She dedicates her activity to advise mainly digital economy players, including many start-ups, in the development of their business, protection and enhancement of their intangible heritage. She has developed a particular expertise in ICT law and innovative technologies, as well as in the field of the valuation of research and public or private data. After having



worked in major international law firms, she founded MetaLaw, a law firm dedicated to the actors of innovation with two other partners (2018). She is also Lecturer at University since 2010 in IP and ICT law. She is currently Lecturer with the Information and Communication Department of the Catholic University of Paris (ICP).

Title: *Legal issue of Big Data: using law as a tool to build a Big Data strategy in the cultural field*

Abstract: Data is the new gold of the twenty-first century: data control has become a key competitive issue for all economic actors. For cultural institutions, the promise of Big Data is based on a better knowledge of audiences to promote a diversification of cultural offerings. In this context, institutions must capitalize on visitor data to build a real development strategy. The Big Data revolution is confronted with the implementation of numerous and complex legal rules (protection of personal data, intellectual property on data mining technologies, ownership of dataset), often experienced by economic actors as heavy constraints.

The presentation will demonstrate that Big Data also pushes the boundaries of law and challenges the role that the law can play in the setting up of an ethical legal system of exploitation of the data beyond the role of the legislator. He is no more intended to be the sole actor of regulation. In this context, it will be emphasized that the law is a vector of innovation enabling the economic actors to appropriate this stake by the modelling of an innovative contractual data governance in the service of building a Big Data strategy

David Haskiya works at the Swedish National Heritage Board. There, he leads a digital team in support of the digital transformation of Swedish museums. The team develops best digital practices, hosts a cultural heritage incubator for small businesses and entrepreneurs, and aggregates and distributes museum collections to Europeana and other platforms. You can find him on Twitter as @DavidHaskiya

Title: *Challenges and opportunities for Swedish museums in digitising their collections*

Abstract: In Sweden it is only a few of the major national museums that could be said to face some Big Data-challenges. There are however museum collaborations in which collections are aggregated where a broader spectrum of Big Data-challenges come into play. In this presentation I will identify the major strategic challenges facing museums in digitising their collections, as individual institutions and at the aggregated/community level. Finally, I will outline opportunities for museums and the museums community to face and overcome these challenges and transform collections data from perceived burden to a valued asset.

.....

A kiwi technology entrepreneur with a background in computer science, **Angie Judge** leads an award winning team at Dexibit, the company she founded to transform decision making in the arts. Angie developed her passion for analytics in the telecoms industry with a corporate career at Hewlett Packard and Amdocs. She is the Chair of the American Alliance of Museum's Technology Board, Chair of the MCN's Data & Insight Group and



host of a tourism analytics think tank network in the United States. In her spare time, Angie mentors young women into technology careers, lectures in data science and is a keen water skier.

<https://www.linkedin.com/in/angiejudge>

solution architecture and organizational leadership, common pitfalls and critical success factors and case study outcomes.

Recent speaking engagements:

- TEDx 2018
- AWS Transformation Day keynote
- Microsoft Ignite keynote
- American Alliance of Museums 2016 - 2018
- Museums and the Web 2016 - 2018
- Museum Computer Network 2017 - 2018
- Shadowtech motivational speaker 2016 - 2018

Title: *"Predicting the future for museums using big data and AI"*

Abstract: Over the past few years, Dexibit has worked with a number of visitor attractions - particularly museums and galleries, of all shapes and sizes from around the world as they have entered an age of big data and artificial intelligence. Using this technology such as machine learning and natural language to predict and analyze visitor behavior, these venues have enabled the democratization of data and powered efficient growth with insight informed decisions. From forecasting visitation to simulating exhibitions to sentiment analysis, Angie will cover the best practices which have emerged from industry in





متحف البيانات الكبيرة

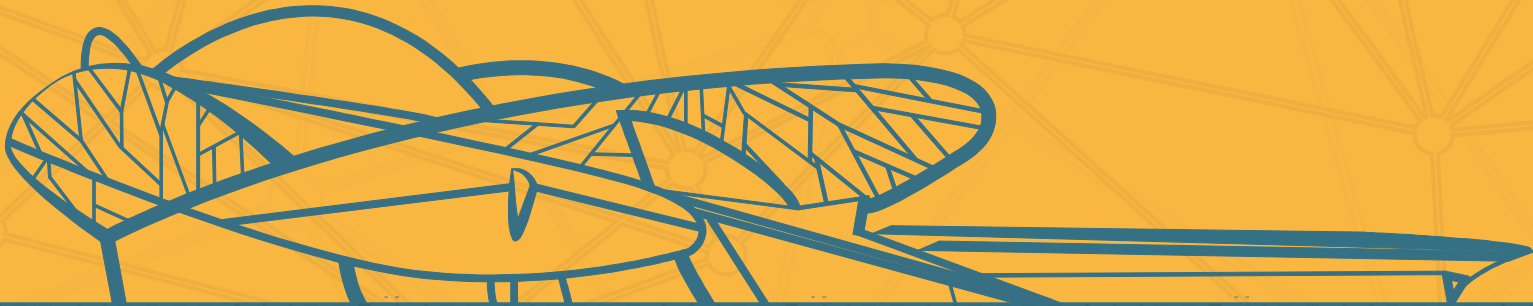
الممارسات والآفاق



MUSEUMS AND BIG DATA

المؤتمر الدولي الأول

٣٠ أبريل - ٢ مايو ٢٠١٩



مكتبة قطر الوطنية، قاعة المحاضرات

٣٠ أبريل

من ٢:٣٠ عصراً إلى ٦:٣٠ مساءً

كلية لندن الجامعية قطر، غرفة 1D02

١-٢ مايو

من ٩:٣٠ صباحاً إلى ٣:٣٠ عصراً

