EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2021)
A. Chalmers and V. Hulusic (Editors)

Bridging the discipline gap: towards improving heritage and computer graphics research collaboration

J. Happa[1,2], T. Bennett[3], S. Gogioso[2], I. Voiculescu[2], D. Howell[4], S. Crawford [3], K. Ulmschneider[3] and C. Ramsey[3].

[1]Information Security Group, Royal Holloway, UK
[2]Department of Computer Science, University of Oxford, UK
[3]School of Archaeology, University of Oxford, UK
[4]West Dean College of Arts and Conservation, UK

Abstract

In this project we are investigating the requirements to ease interdisciplinary collaboration between computer graphics researchers and heritage-related researchers who work with shared graphics-related datasets. We postulate that most challenges can be overcome by ensuring that datasets (irrespective of discipline) are captured, processed and disseminated in ways that accommodate the needs of as many disciplines as possible – making the datasets more useful and more usable. This is not to say that a union of all discipline methodologies is required, but instead: we deem it necessary to identify what changes are feasible in existing (discipline-centric) practices to maximise the benefits, while limiting resource costs. The purpose of this paper is to begin this conversation, present our project, preliminary results and where the project will go next. We also propose the outline of an interdisciplinary, peer-reviewing framework that can be used across disciplines.

CCS Concepts

• ***Computing methodologies*** → ***Computer graphics;*** • ***General and reference*** → ***Document types;***

1. Introduction

Different research disciplines have dissimilar practices and priorities with regards to data capture, storage, analysis and display of real-world measurements. In this project we are investigating the requirements to ease interdisciplinary collaboration between computer graphics researchers and the wide range of heritage researchers with regards to capturing, appending metadata, processing, and disseminating graphics-related datasets. We argue that research should adopt principles that maximise utility of datasets and procedures for as many disciplines as possible. We postulate that most challenges related to interdisciplinary collaboration can be overcome by ensuring that datasets (irrespective of discipline) are captured, processed and disseminated in ways that accommodate the needs of as many disciplines as possible – making the datasets more valuable for research purposes across disciplines. We are investigating how existing formats and standards are used; whether best-practices from various disciplines are compatible with each other and whether an inclusive/combined best-practice is practicable. We conducted a feasibility study and a preliminary questionnaire in various disciplines to explore these problems further. This paper presents our project, our preliminary results and where the project will go next. We lay the groundwork for an interdisciplinary, peer-reviewing framework that can be used across projects that makes use of computer graphics to study cultural heritage.

2. Related Work

Several formats, standards and guidelines exist to study heritage using computer graphics. Perhaps two of the most widely known guidelines are the London Charter [BDN*06, CGTFA*13] and the Seville Principles [LMG11, Ben13, CGTFA*13]. Both act as useful starting points, but focus on computer-based visualizations, whereas we are examining the requirements for collaboration across graphics and heritage-related disciplines. The CIDOC Conceptual Reference Model (CRM) [Hun02, Doe05, BMT08], Functional Requirements for Bibliographic Records (FRBR) [Til05] and Dublin Core [WKLW98] are often listed as examples of frameworks available for categorization of data in projects. These have been built with specific users and act as foundational reference points. The FAIR principles [WDA*16] for scientific data management and stewardship help researchers improve data *Findability, Accessibility, Interoperability* and *Reuse* of digital assets. While these principles are also a good first step, they do not propose how to tackle problems related to longevity of datasets (i.e. ensuring that data is both relevant and valuable (in research) today and in the foreseeable future). Another problem is the R1.3 reuse principle that specifies that "*data meet domain-relevant community standards*". This has the potential to allow best-practices in one discipline to be in direct odds with another discipline, even if they make use of the same data. Instead, data should meet a scientific benchmark that is valid across as many disciplines as possible.

3. Project Description

Advances in our understanding of cultural heritage are often driven by interdisciplinary collaboration and new modes of analysis drawing upon the integration and correlation of different expert methodologies and forms of study. Yet currently, there is no unified, structured framework to bring to bear all the resources available in disparate locations and disciplines on significant questions of material culture. The analytical tools and expertise among the authors of this paper are divided among several disciplines e.g. heritage science, archaeology, conservation science and computer graphics (incl. geometric modelling and physically-based rendering). Our working group saw an opportunity to establish a framework to facilitate interdisciplinary collaboration in cultural heritage and computer graphics research. Each member in the group have contributed at least one dataset (from a past or current research project), ranging from hyperspectral images, to reflectance transformation images to traditional photography images, that they believe has value across disciplines, and with improvement-suggestions via peer-reviewing we expect to identify differences in discipline practices. The working group have worked on developing a framework for data sharing and peer review. In this phase of the project, we have identified a preliminary set of requirements for interdisciplinary collaboration. In the future, we expect to study factors that can negatively and positively impact interdisciplinary collaboration.

3.1. Objectives

The core objectives of this project are the following:

- **Methodological Objective**. Development of a peer-reviewing framework for a suite of templating options, principles and good practices for researchers and practitioners.
- **Theoretical Objective 1**. Identify how gaps in discipline standards hinder straightforward interdisciplinary collaboration.
- **Theoretical Objective 2**. Develop a foundation to enable easy sharing and analysis of graphics content across different disciplines, in particular: measurement, assessment, improvement and communication between domain experts.
- **Empirical Objective 1**. Collect insight about how real-world domain experts of different disciplines use data and compare commonalities and dissimilarities in procedures and dataset collection and maintenance.
- **Empirical Objective 2**. Evaluate current practices and propose how they can be improved through peer-review. Identify critical factors that affect performance, including usability and functional relationships.
- **Tool-Development Objective**. Document and iteratively improve approaches to measure, evaluate, improve datasets and tools for interdisciplinary sharing, collaboration and analysis, including methods for better annotation of domain expert insights to facilitate interdisciplinary communication for iterative improvements.

3.2. Feasibility Study

Members in our working group contributed datasets or analysis on datasets that they believe have value across disciplines. Through peer-review and informal discussions, challenges for collaboration have been and are still being identified. Suggestions are being made to tackle challenges through peer-reviewing of other datasets and procedures. The datasets we used are from current and past projects of the authors of this paper. Datasets and procedures in question include: acquisition and development of 3D virtual environment data (from 3D modelling tools with geometry, texture maps and Bidirectional Reflectance Distribution Functions (BRDFs)); Hyperspectral Imaging and; Reflectance Transformation Imaging (RTI) data. Through our discussions, we are proposing a peer-review framework, as shown in Figure 1. It is comprised of five main steps:

- **Assembly**: gather stakeholders, relevant datasets and known practices/procedures about how those datasets were collected;
- **Calibration**: set expectations and scope of the collaboration, but also identify knowledge gaps and fill these in by using a shared vocabulary;
- **Peer review**: each stakeholder review other stakeholders' datasets and procedures to find ways of improving datasets and procedures for their discipline – this could be done as individual discipline feedback and be a catalyst for group discussions;
- **Changes to data (and procedures)**: adopt the suggested changes from the peer-reviewing stage, including for data collection, data processing, data storage and data dissemination (including data sharing);
- **Validation**: demonstrate collaboration improvements.

4. Questionnaire

We used our informal discussions from our study as a starting point inform us about what questions may be useful to ask individual researchers about their graphics and heritage collaboration experiences. The questions are listed in the Appendix. Here, we detail the purpose of each question (design) and summarise the results. We used convenience sampling to 12 selected relevant experts - from within and outside our working group. The questionnaire acts as a pilot study. In the future, and based on the results from this pilot questionnaire, we expect to create and distribute a revised version to a broader range of academics and practitioners. The purpose of this questionnaire is to obtain some preliminary responses from known entities who have a background in the field.

Q1 was used as a control to ensure participants were able to answer the questions at hand. Q2 was intended to identify the distribution of disciplines, while Q3 identified how long each expert has worked in their respective discipline. For Q2, 7 participants were from computer graphics, while 1 participant were each from Digital Humanities, Archaeological Archives and Conservation. Finally, 2 participants were from Archaeology. For Q3, 6 participants have 20+ years of experience in their discipline. 3 participants have 15-20 years of experience, 2 participants had 10-15 years experience and 1 participant had 5-10 years of experience.

Q4 was used to identify the types of graphics content used by the experts. 9 participants stated they work with 3D models, all 12 work with common image types, 8 work with specialised images such as RTI or multispectral/hyperspectral imaging, 5 work with multiple measurements of a single point, 5 work with single (or averaged) measurements of a point. 1 participant added an "Other" (morphing of images of archaeological sites; geographical coordinates). Q5 then asked participants to list their main types of graphics content. 8 participants listed imaging, while 4 listed 3D models as their primary graphics content.

Q6 identified commonly used standards and software for their projects. 3 participants specified uses of both metadata standards and metadata software. 11 participants made use of common imaging formats, but only 2 made use of imaging standards and guidelines, 6 participants make use of open Source calibration and processing software, 3 participants make use of online calibration and

Figure 1: *Preliminary peer-reviewing framework*

processing software, 10 participants make use of proprietary calibration and processing software, and finally 1 participant listed an "Other", which they stated to be an open source digital asset management system. Similar to Q5, Q7 asked participants to list the main ones they used, these include: Blender (x2); Photoshop (x3); Meshlab (x2); Maya (x2), Matlab (x2); CHI DLN; DNG; Metashape; 3D modelling tools and proprietary analysis tools; Resource space; Epson scan.

In Q8, 9 participants stated that they have developed one or more pieces of software to meet a goal or capability that existing software did not provide, while 3 participants said this was not the case for them. The missing capabilities were listed in Q9, including: *Matlab for batch-processing RTI datasets; new algorithms (x2); step counting; HDR capabilities; keeping track of process history and to validate image sets followed correct rules, such as no sharpening, tone curves, camera settings, etc.; recovering lighting from images; 3D scene viewer for dissemination purposes this was before unreal and unity were so readily available. BRDF visualizer – none exist; crowd-sourcing image meta-data information (i.e. archaeological site); re-photography mobile phone app with geolocation app*. We make note that to our knowledge there is at least one web-based BRDF viewer available [JVAP*21].

In Q10, we asked whether the participants have experienced challenges when working with graphics-related datasets and formats in their discipline. 10 participants said yes, 2 said no. Q11 asked participants to highlight what those challenges were. These included: *Producing linear raw files using Matlab and compiling DCRAW for Macs; Inconsistent formats, lack of physically-based rendering support; missing bits; missing datasets; Conversions are still tricky; BRDF datasets tend not to be homogenous; Assessing digital representations, especially 3D models, without knowing what was done to the models - such as smoothing, hole filling, hand editing, etc. This data is often not available or not published with the model; Data sets are often not interoperable, materials appear visually different depending on renderer used; Conversion between formats isn't always lossless. While many principles exist for good practice, none are detailed enough for specifics of concrete data formats - either they are too conceptual in nature or haven't thought of my use cases.; transferring data; different standards; exporting and importing metadata and images keywording and search related issues; uploading and downloading issues when trying to link crowd-sources data / mobile app data with database, costs and compatibility issues; long-term data storage issues*.

In Q12, we identify challenges when collaborating across disciplines w.r.t. graphics and heritage. Similarly to Q10, 10 participants said yes, 2 said no. Q13 asked participants to idenify these challenges. These includes: *Dissemination and annotation of high-resolution RTIs / 3D models over the web; use of high-resolution 3D models in AR/VR formats; Different standards; sharing; Terminology and data conversion; Just CAD conversions are sometimes tricky, e.g. CATIA to Maya.; It's hard to know what's been done to produce digital representations. Systems such as sketchfab don't provide tools to make this easy. There isn't agreement among researchers in heritage – let alone with folks outside of heritage about what data to record and how to manage these issues; Converting between proprietary data formats and data formats for tools that are free or open; Comparing different techniques; Language barriers – across disciplines the same words can have different meanings e.g. artifact or texture; Different disciplines may have different data analysis requirements. E.g. colour calibrated images are usually necessary in graphics and archaeology, but may not be the case for some art historians; transferring data; different standards; exporting and importing metadata and images; keywording for other disciplines - no standards*.

Q14 asked participants to identify key data storage and data sharing issues they have experienced. 10 agreed on *Data cannot be shared easily because of data volume (e.g. number or files or size of files)*; 8 agreed on *Data cannot be shared easily because of content formats across disciplines (e.g. incompatible image formats)* and *Data formats becoming obsolete over time (and thus unusable or having to convert them)*; 7 agreed on *Losing metadata while creating backups or recovering data* and *Losing data due to mistakes*; 6 agreed on *Data not being 'as good' as when it was first captured because technology keeps improving (i.e. the usefulness of the data decreases over time)* and *Data or metadata not detailed enough to contain all necessary content (e.g. calibration parameters not recorded or annotations not recorded easily or similar)*; 5 agreed on *Lack of quality-verification tools (e.g. inability to check integrity of data or check data was captured using an appropriate calibration or similar)*; 3 agreed on *Data theft (either digital copies or physical theft)* and *Losing accuracy of data during storage, copy or transmission of data*. One participant stated *Standardised automated analysis capabilities*. One participant stated *system updates creating problems; update and security leading to issues with need resolving; fragility of electronic record over analogue archives*. Finally, one participant stated *none*.

Q15 asked how we can improve data longevity (i.e. usefulness and integrity). The answers provided were: *The CIDOC CRM is difficult to learn and apply, however CHI's Digital Lab Notebook software helps to organize relevant info without needing to learn the details of the CRM. It also helps to record metadata related to calibration and image capture geometry; github, or any open source repo; CSV; Use universal formats; Provide high quality metadata about capture and processing. Choose archival formats for image data - such as DNG. Save all the original image data, store images, metadata and digital representations in repositories such as based on Fedora. It's most useful if repositories are readily available such as the Archaeology Data Service, and also through university libraries; Share data more widely, many researchers publish their new techniques but don't share their data or their code. Better standards for data exchange across different tools; Document decision made throughout the project, make use of reasonable best of breed technologies that have transparent data processing and storage capabilities, i.e. the way data is captured, store and processed should not be a secret hidden away by the*

tools used. This is necessary to ensure reproducibility; need excellent IT support; funding for servers to store data in different (physical) locations; only way to be sure is to keep printed out record as we cannot be sure that updates do not loose [sic] data.

Q16 asked participants to state any other challenges they have faced. 6 participants raised issues, including: *I've had problems with image time-stamps being out-of-sync between the camera, lab computer (not connected to internet), and data storage and backups; in some cases the heritage world should embrace more technology; It is difficult to find a balance between the need for high quality metadata, saving the original empirical data, and the costs and time associated with doing it well; The 'all the research problems in this area have been solved' myth.; It's still a significant undertaking to reconstruct a heritage site well; linking databases of images constant stream of technology with few standards; how can we transfer to those standards, (i.e. IIIF, Palagios etc); costs of transfer.*

In Q17, no concerns were raised about the design of this questionnaire, but one participant stated that collecting information about this often overlooked topic is appreciated.

5. Conclusion

In this paper, we have presented the preliminary work of a small working group on how to improve collaborations between graphics and heritage researchers. We continue to investigate how different research disciplines have dissimilar practices and priorities with regards to data capture, processing and dissemination of data. We proposed a interdisciplinary, peer-reviewing framework that could be used across disciplines in studies of historical content. Our pilot study (questionnaire) was limited in scope, but identified a number of concerns. There is a large debate in the research community on standardization in the archaeology and virtual archaeology domains, in particular on documentation, analysis, maintenance and dissemination of heritage data. We believe research field remains fragmented, and to enhance the interdisciplinary value of datasets, we deem it necessary to increase data longevity and improve collaboration capabilities of researchers, by sharing use-cases across disciplines and applying lessons learnt on datasets in question. To those experienced in the topic of computer graphics and heritage, many of our findings will be predictable and will need deeper discussions. The purpose of this work is to provide empirical evidence of this common knowledge, and provide motivation for future research and conversations.

Future work includes expanding our trials of data sharing tools such as detailed case studies, wikis, version control systems, slack, gitlab, server collaboration solutions and identify usability considerations, knowledge gaps in the community and comparison of best-practice procedures across disciplines. Further studies, include an expanded questionnaire with a wider scope and significantly more data points. Our project is still in its infancy. We aim promote a novel type of collaborative, interdisciplinary framework, which we hope may eventually lead into a laboratory offering for real-world measurements and procedures where researchers can request samples to be collected, processed and analysed – as well as easy means to generate new collaboration templates for heritage and graphics research collaboration. We welcome readers to reach out to us for further discussions.

References

[BDN*06] BEACHAM R., DENARD H., NICCOLUCCI F., ET AL.: An introduction to the london charter. *Joint Event CIPA/VAST/EG/EuroMed* (2006). 1

[Ben13] BENDICHO V. M. L.-M.: International guidelines for virtual archaeology: The seville principles. In *Good practice in archaeological diagnostics*. Springer, 2013, pp. 269–283. 1

[BMT08] BINDING C., MAY K., TUDHOPE D.: Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM. In *International Conference on Theory and Practice of Digital Libraries* (2008), Springer, pp. 280–290. 1

[CGTFA*13] CARRILLO GEA J. M., TOVAL A., FERNÁNDEZ ALEMÁN J. L., NICOLÁS J., FLORES M.: The london charter and the seville principles as sources of requirements for e-archaeology systems development purposes. *Virtual Archaeology Review 4*, 9 (2013), 205–211. 1

[Doe05] DOERR M.: The CIDOC CRM, an ontological approach to schema heterogeneity. In *Dagstuhl Seminar Proceedings* (2005), Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 1

[Hun02] HUNTER J.: Combining the CIDOC CRM and MPEG-7 to describe multimedia in museums. 1

[JVAP*21] JASPE-VILLANUEVA A., AHSAN M., PINTUS R., GIACHETTI A., MARTON F., GOBBETTI E.: Web-based exploration of annotated multi-layered relightable image models. *Journal on Computing and Cultural Heritage (JOCCH) 14*, 2 (2021), 1–29. 3

[LMG11] LOPEZ-MENCHERO V. M., GRANDE A.: The principles of the seville charter. In *CIPA symposium proceedings* (2011), vol. 2011, pp. 2–6. 1

[Til05] TILLETT B.: What is FRBR? a conceptual model for the bibliographic universe. *The Australian Library Journal 54*, 1 (2005), 24–30. 1

[WDA*16] WILKINSON M. D., DUMONTIER M., AALBERSBERG I. J., APPLETON G., AXTON M., BAAK A., BLOMBERG N., BOITEN J.-W., DA SILVA SANTOS L. B., BOURNE P. E., ET AL.: The FAIR guiding principles for scientific data management and stewardship. *Scientific data 3*, 1 (2016), 1–9. 1

[WKLW98] WEIBEL S., KUNZE J., LAGOZE C., WOLF M.: Dublin core metadata for resource discovery. *Internet Engineering Task Force RFC 2413*, 222 (1998), 132. 1

Appendix

Below are the questions in the questionnaire as they were asked.

- Q1 Have you used computer graphics (e.g. images, imaging techniques, 3d models, virtual/augmented reality, graphics APIs or other) for one or more heritage-related projects (select one)?
- Q2 What is your main field of research/discipline (select one)?
- Q3 How long have you been in your discipline (select one)?
- Q4 What type of graphics content have you worked with (tick all that apply)?
- Q5 Please list the main ones you use in your research:
- Q6 Do you use any standards and software for any of your projects (tick all that apply)?
- Q7 Please list the main ones you use:
- Q8 Have you (or someone on your team) developed one or more software for your project to meet a particular goal or capability that existing software did not provide (select one)? Q9 If yes, can you describe this goal or capability was?
- Q10 Have you experienced challenges when WORKING WITH graphics-related datasets and formats in your discipline? (i.e. do existing graphics software, formats and standards cause you theoretical or practical issues) (select one). Q11 If yes, what were those challenges?
- Q12 Have you experienced challenges when COLLABORATING WITH other disciplines than your own on topics related to graphics and heritage? (i.e. do existing graphics software, formats and standards cause you theoretical or practical "sharing of data across disciplines or digital systems" issues) (select one). Q13 If yes, what were those challenges?
- Q14 Have you had any of the following data storage and data sharing issues (tick all that apply)?
- Q15 How can researchers feasibly ensure that data can last for as long as possible: in terms of usefulness (between research disciplines), data integrity preservation and data access (optional)?
- Q16 Are there any other challenges related to graphics and heritage you would like to flag (optional)?
- Q17 Have you had any issues with the design of this questionnaire? Please raise them here (optional).