Born Digital, Born Slippery

Data Modelling for Born Digital/Hybrid objects:
March 2022

Born Digital, Born Slippery
Data Modelling for Born Digital/Hybrid objects

Author

Richard Palmer, Victoria & Albert Museum

Acknowledgements

Lead research organistion

In partnership with

Funded by

Contents

1. Introduction

The report *Preserving and sharing born-digital and hybrid objects from and across the National Collection* discusses the challenges for cultural heritage institutions in cataloguing, preserving, and providing access to four born digital and hybrid objects/projects. This response looks at the issues brought up by those discussions, in particular for the purpose of cataloguing, and the extent, within existing cataloguing standards, they can be supported. Where there does not appear to be support within current standards, model development activity is proposed to allow the expansion of the current `born physical` focused cataloguing standards to also handle born digital and hybrid object cataloguing.

Background

Cultural heritage cataloguing has evolved to suit a variety of different needs, from managing inventory to tracking loans to documenting object histories. Documentation choices made at each institution reflect their individual priorities and interests in the history recorded; enabled through the construction of a model that describes the physical object through the uses of data which seeks to describe its reality. Due to these locally made choices, cultural heritage institutions are unable to easily exchange their data (as one model does not simply match across to another), and researchers face a similar 'interoperability' challenge when trying to bring this data together from multiple institutions (for example to examine the data on all the objects made by one artist or maker held across multiple collections). This challenge has been long recognised by the sector, and attempts have been made to develop interoperable data models that can be used to share data between institutions (either by using this data model

locally for "data at rest" in core systems such as a collections management system, or to map (as much as is possible) between the local data model and the interoperable data model). These standards can be in the form of guidance such as the Spectrum [1] standard developed by Collections Trust [2] which recommends procedures and types of information that should be recorded (but does not enforce a particular structure, so does not resolve the issue of interoperability) or it can be in more formalised data models (aka ontologies) such as CIDOC CRM [3] which enforce a logical structure. More recently, a consortium of cultural heritage institutions started developing the Linked Art [4] profile of CIDOC CRM aiming to ease the route for institutions planning to implement (and for research users [5] to make use of) such an interoperable model.

The development of the existing data models such as those mentioned above have been predicated on the legal acquisition by the cultural heritage institution of a physical object. The introduction of born digital/hybrid collecting proves problematic as this core assumption associated with physical objects (e.g., fields in the data model assuming the object has a physical form) do not always apply, or allow for comprehensive cataloguing of the born digital object(s) (e.g., controlled vocabularies such as Getty's AAT [6] to define materials and techniques are drawn from physical objects). This leads to some of the workarounds as described in the report or from existing born digital objects in the V&A's collection, such as:

- a physical object is artificially used to represent the whole project (for example a computer game documenting the physical cartridge it is stored on [7])

---

[1] https://collectionstrust.org.uk/spectrum/
[2] https://collectionstrust.org.uk/
[3] https://cidoc-crm.org/
[4] https://linked.art/
[5] E.g For uses as documented in 'Collections As Data' https://collectionsasdata.github.io/
[6] https://www.getty.edu/research/tools/vocabularies/aat/

- digital objects are recorded as a representation of a physical object rather than an object in themselves

These cataloguing workarounds can then lead to confusion for those accessing a collection as to what the catalogued object is, and what the distinction is between the object and the digital representation of it (i.e., a digital image of a pipe is not the object [8], but if the object acquired is a born digital image, does it also have a separate representation?).

To resolve this situation, where the data model does not satisfactorily capture some aspect of reality it is modelling, data models can be extended by the gathering of example objects that indicate what additions and modifications are needed to fill these modelling gaps. The case studies for Born Digital/Hybrid serve as a sample that can be used to sketch out this work.
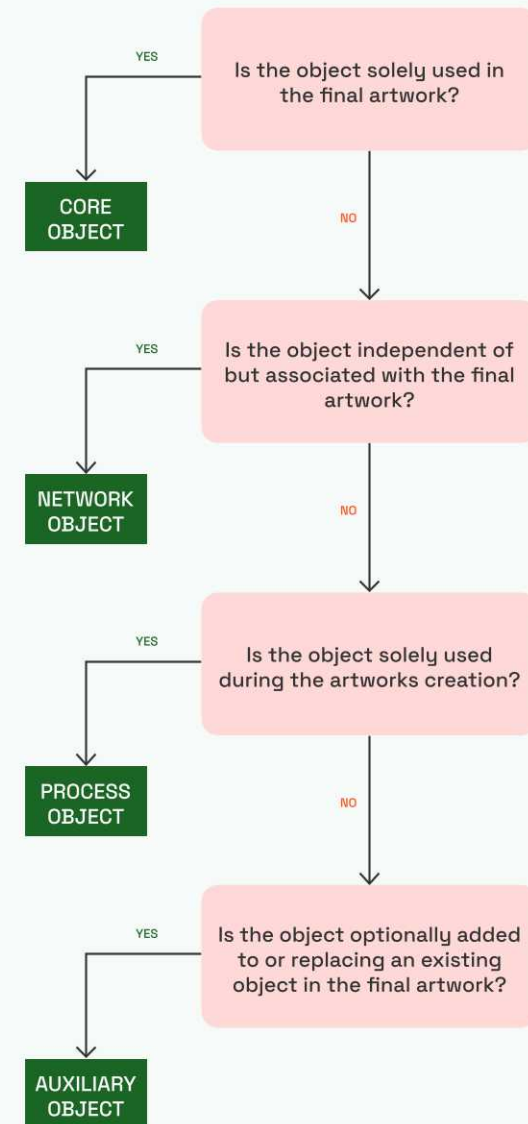
[7] https://collections.vam.ac.uk/item/0107014/sonic-the-hedgehog-computer-game-sega/
[8] See https://collections.lacma.org/node/239578

2. Case Studies

A summary of each case study follows, with examples of different forms of cataloguing identified in each case. As there are many commonalities of form between the examples, these are grouped as below, and the relevant type indicated for the objects in each case study; these groupings are then explored in more depth in Modelling Requirements. Within the core object, the project or conceptual project are understood as the culturally received idea of the object before any attempt at modelling it, and before identifying its associated parts, versions and processes.

- **CORE** object(s) – Cataloguing an object/project and its (core) digital/physical parts

- **NETWORK** object(s) – Cataloguing the network of objects that are associated with the core project but can exist independently

- **PROCESS** object(s) – Cataloguing the process objects created as intermediate steps of the core object

- **AUXILIARY** object(s) – Cataloguing the auxiliary objects that can be used for different versions of the core object [9]

- **WORKFLOW** – Cataloguing workflows both from the creation process and the conservation processes

- **RECEPTION** – Cataloguing the audience reception of the core object(s)

- **COMMUNITY** – Cataloguing multiple viewpoints on the core object(s) from beyond the collecting institution

Is the object solely used in the final artwork?

— YES → CORE OBJECT

— NO →

Is the object independent of but associated with the final artwork?

— YES → NETWORK OBJECT

— NO →

Is the object solely used during the artworks creation?

— YES → PROCESS OBJECT

— NO →

Is the object optionally added to or replacing an existing object in the final artwork?

— YES → AUXILIARY OBJECT

---

[9] This has similarities to the approach discussed for cataloguing performance artworks in allowing some elements to be constant and some to be in flux. See '*Developing a strategy for the conservation of performance-based artworks at Tate*'
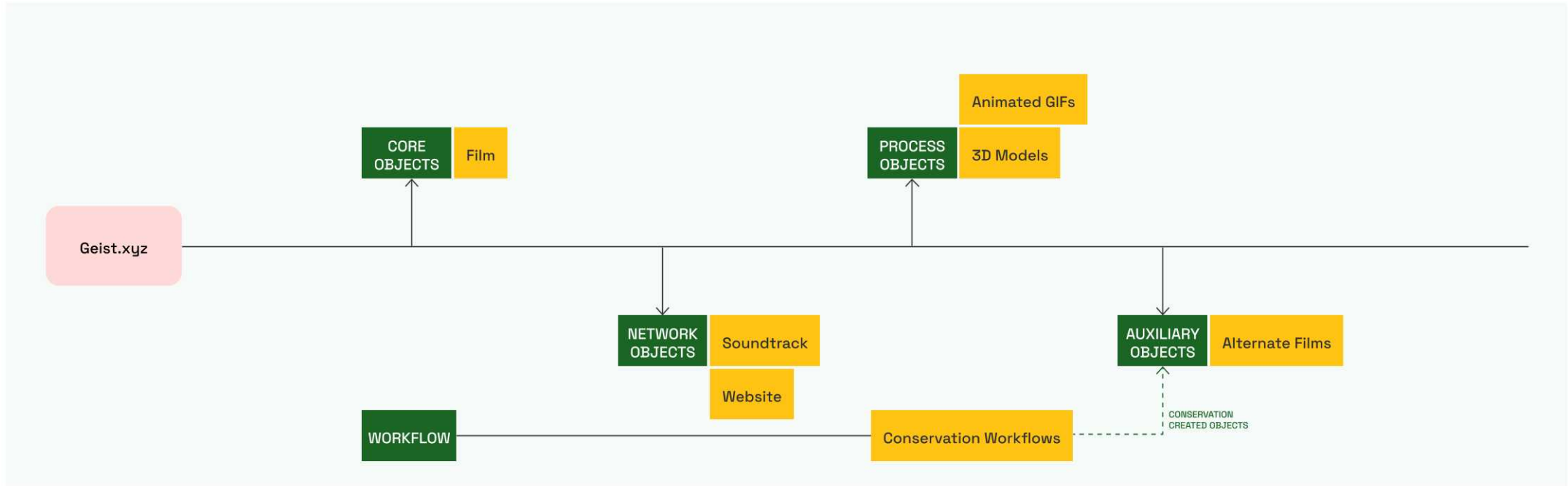
Geist.xyz

'geist.xyz' is 'an experimental film created using computer generated imaging techniques, most often seen in videogames, to manifest '0% organic' textiles and textures moving across a gravity defying virtual environment'. For accessioning the project to the V&A collection, the following cataloguing requirements were identified:

- **CORE** – The project 'geist.xyz' and its digital part (a special edition of the film)

- **NETWORK** – The soundtrack (which has both digital and physical parts), the website

- **PROCESS** – The animated .gifs and 3D objects created during the development process

- **AUXILIARY** – Alternate film versions

- **WORKFLOW** – Cataloguing the different processes and workflows involving all the above objects involved in creating the objects (e.g. 3D render workflows)

In the Eyes of the Animal
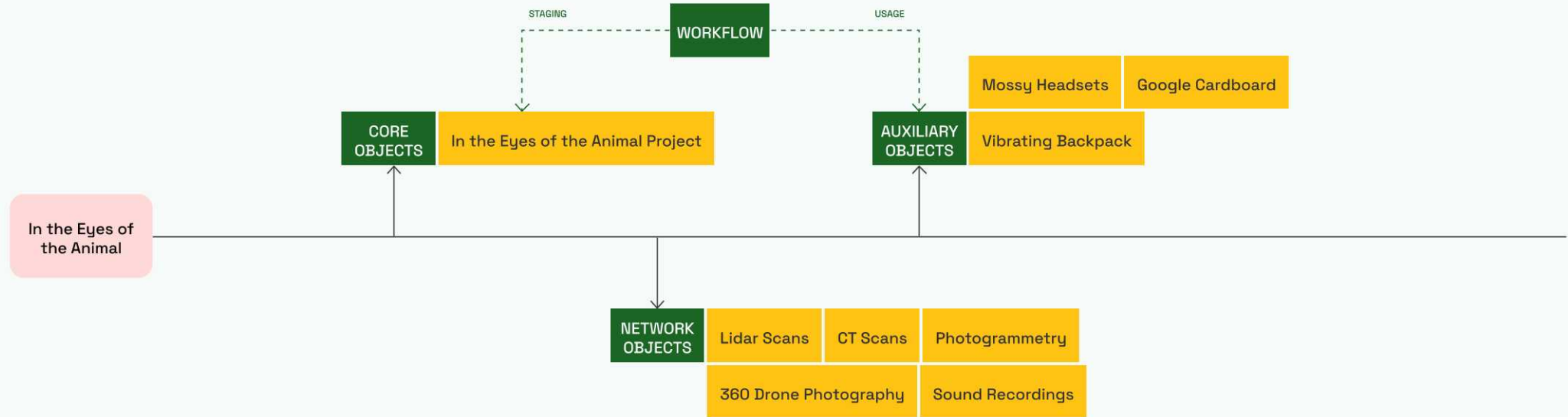
'In the Eyes of the Animal' is 'a Virtual Reality project, offering users the sensory exploration of a forest from the perspective of three different animals: dragonfly, frog and owl'. For accessioning the project to the V&A collection, the following cataloguing requirements were identified:

- **CORE** – 'In the Eyes of the Animal' project concept

- **NETWORK** – Lidar scans, 360-degree drone cameras, computerized tomography, photogrammetry, sound recordings

- **AUXILIARY** – Mossy Headset, vibrating backpack, Google Cardboard

- **WORKFLOW** – Modifications made for performances in different spaces

- **RECEPTION** – 360 video capture of VR experience

Y-Stop
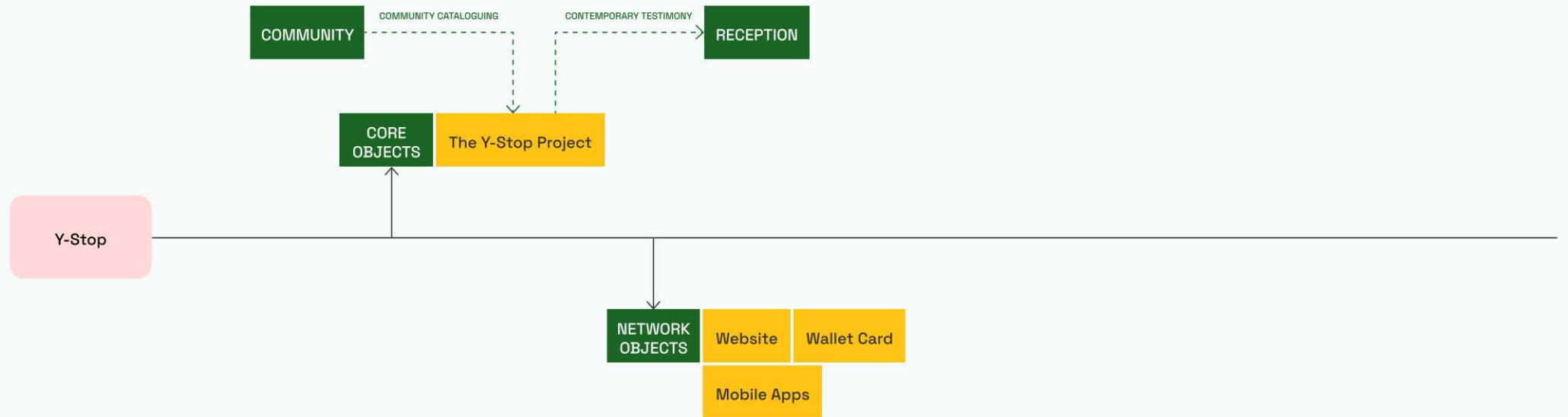
Y-Stop is 'a website, a mobile app and a set of print and digital informative materials to guide people (especially young men from minority backgrounds) in managing a stop and search interaction and in potentially seeking legal support'. For accessioning the project to the V&A collection, the following cataloguing requirements were identified:

- **CORE** – The Y-Stop project (ongoing)

- **NETWORK** – Website, Mobile App (ios/Android), Wallet Card

- **RECEPTION** – Outreach work, Contemporary testimony

Instagram
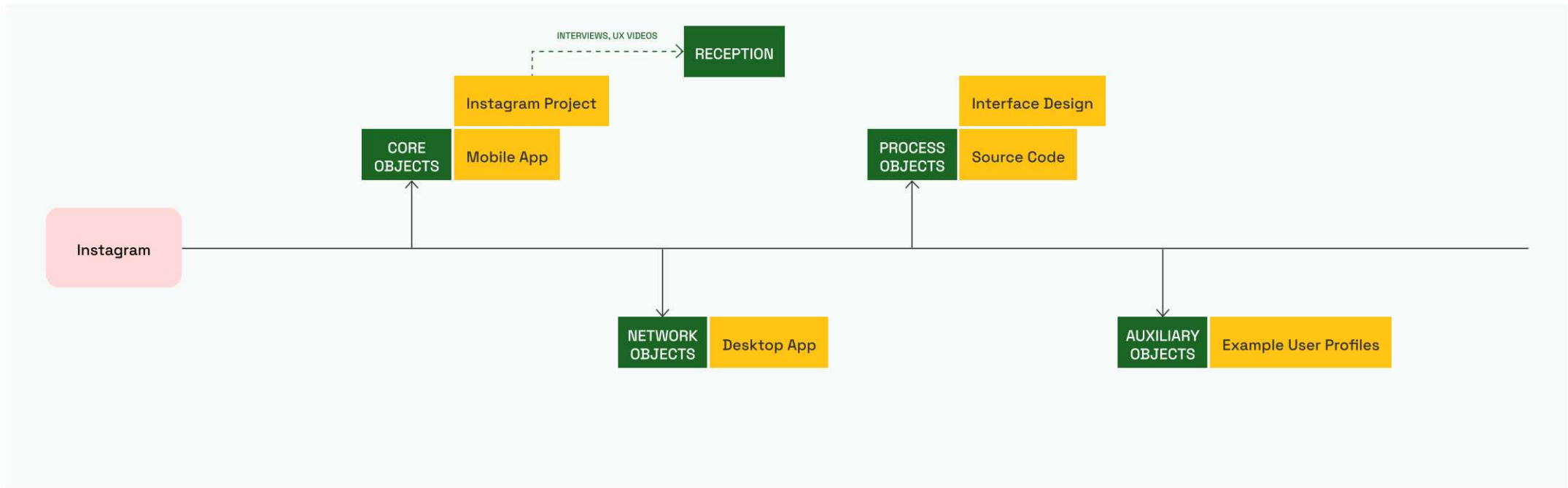
Instagram is 'a photo and video sharing social media platform released in 2010 as an iOS app only, with an Android and a desktop version released in 2012, when the service was acquired by Facebook Inc (now Meta)'. Proposed objects to be catalogued for a hypothetical acquisition:

- **CORE** – The Instagram project (ongoing), Instagram App
- **NETWORK** – Desktop Application
- **PROCESS** – Interface Design, Source Code
- **AUXILIARY** – Example User Profiles
- **RECEPTION** – Interviews, Videos of User Experience

3. Modelling Requirements

Taking the functionality required from the case studies above, we can examine the existing CIDOC CRM data model and whether it can provide this functionality. Before we inspect those, a brief review of how a data model such as CRM is specified. A data model sets out to define:

"the underlying semantics of database schemata and structured documents used in the documentation of cultural heritage and scientific activities. In particular it defines the semantics related to the study of the past and current state of our world, as it is characteristic for museums, but also or other cultural heritage institutions and disciplines. It does not define any of the terminology appearing typically as data in the respective data structures; it foresees, however, the characteristic relationships for its use. It does not aim at proposing what cultural heritage institutions should document. Rather, it explains the logic of what they actually currently document, and thereby enables semantic interoperability." [10]

It does this through the uses of two data modelling 'primitives':

- **Classes** – A category of items sharing one or more common traits which can be used to identify items belonging to the class (for example – "Human-Made Object" is a class that groups together all things that satisfy the criteria of being a human-made object)

- **Properties** – A property exists to define a relationship of a specific kind between two classes (for example "depicts" allows one class to indicate it depicts another class)

A select group of classes are defined in the core standard intended to enable broad support for modelling all types of cultural heritage objects. Extensions to these core classes are then available, providing additional classes and properties developed for richer modelling support for particular usages, for example CRMba [11] for modelling architectural buildings, CRMtex [12] for the study of ancient texts, etc [13]).

Following on from the review of the case studies and the functionality required (as classified into **CORE**, **NETWORK**, **PROCESS**, etc) we can see how these could fit into existing CRM classes and properties, or if they may require further development to support them.

Note:
References in bold starting with **E** (e.g., **E95**) are a conventional short form to refer to a particular class in CIDOC CRM, for example **E95** is the "Spacetime Primitive" class. Likewise, Properties are referred to using **P** prefix (e.g., **P179** for "had sales price"). Classes starting with **D** prefix are from the CRMdig extension (see below).

---

[10] Definition of the CIDOC Conceptual Reference Model Version 7.2, pg. 9
[11] *https://cidoc-crm.org/crmba/*
[12] *http://www.cidoc-crm.org/crmtex*

[13] See *https://cidoc-crm.org/collaborations* for all extensions in development

Digital (Information) Object

A self-evident need from the case studies is to be able to catalogue a born digital object without workarounds (such as cataloguing proxy physical carriers as the object). The existing class for objects (**E22** – Human Made Object) does not allow for objects that do not have a physical form, so continued use of this would simply replicate that workaround. The development of the CRMdig extension to provide digital objects (**D1**) provides a resolution to some of these requirements to allow a digital object (and its creation workflow) to be modelled, but as this class is also used to catalogue the digital representation of a physical object (e.g. a digital photograph taken of the physical object), this usage would be similar in that it represents the contents of the file (e.g. the binary data for the app) and not the object itself. That means some other class (possibly **E73** – Information Object) would be needed in addition, to represent the object, which can then have **D1** (or even multiple **D1**) classes attached to it (e.g., for recording different released versions of the app).

Core Objects

The core objects(s) are those that are considered to be the object/project(s) at the time of acquisition by the institution, these can then be related to other objects that need to be documented in the same acquisition. For a physical object this is handled by the existing class Human-Made Object (**E22**), which can have parts (other **E22** objects) which may appear to immediately provide a solution for hybrid objects (e.g., the Y-Stop project) to catalogue both its physical and digital objects as the core objects, but this appears to be ruled out by the requirement from CRM for all the objects to have physical form. (Other existing ways to record aggregations of objects such as **E78** (Curated Holding) are not intended for use with "collective objects" which would also appear to rule out their use for hybrid objects.)

That leads to the need for a new "binding object" class representing the object/project [14], which can serve to bring together the physical (**E22**) and/or digital object parts (a class also to be developed, see above). Whether this binding could be handled based on an existing core class such as Thing (**E70**) or Conceptual Object (**E28**) or alternatively by use of the LRMoo/FRBRoo [15] extension (based on the FRBR [16] bibliographic data standard to allow modelling of a Work and its various Manifestations) this would need to be the subject of further standard development, including examining the legal and collection reporting consequences of creating this class (i.e., how many objects does a museum acquire when it acquires a hybrid object). There is also the possibility this could satisfy modelling needs beyond born digital objects, there would appear to be scope for cataloguing performance artworks using a similar approach.

Object Relationships

The other types of objects (**NETWORK**, **PROCESS**, **AUXILARY**) can be catalogued within the existing model (once the above necessary classes are available as well), but their relationships to the other objects in the acquisition may need the addition of properties to make clear their collection status (for example, that an auxiliary object can be used as an alternative to an core object, or that a process object is created as part of the work leading up to a core object). There are also legal and reporting considerations, for example, should alternative versions of an object/project be considered as a separate object/project in the collection; and for access purposes, consideration would need to be given as to how these relationships are shown to make clear what is the object/project and what are the objects related to it.

Both **PROCESS** and **AUXILIARY** object types have a strong connection to **WORKFLOWS** as they are involved either in the creation process (**PROCESS** objects) or in the conservation process (**AUXILARY** objects).

---

14   object/project is used here to refer to what might be alternatively called the completed artefact or concept at the time of acquistion

15   https://cidoc-crm.org/frbroo/
16   https://www.loc.gov/cds/downloads/FRBR.PDF

Workflows

The ability to record the activities of the creation process of digital objects has been considered in the CRMdig extension (primarily designed for digitisation workflows). This could be used for recording creative process workflows (e.g., documenting images turned into 3D models turned into films), although extensions to it to capture a wider range of digital creativity would be needed (see for example experimental work on cataloguing born digital 3D models using CRMdig and CIDOC CRM [17]).

A related workflow is needed to record the activity of object conservation, the digital preservation actions that can be taken on the digital object, potentially by creating new auxiliary objects when an emulation of the original object is created for digital preservation purposes. This has the potential to tie in with the idea from performance cataloguing of activation reports [18] to record the staging of a performance-based artwork, in this case the conservation workflow would be carried out on the 'binding object' rather than on the physical or digital object.

Reception

The ability to document the reception of an object/project is something that fits into the work on the CRMsoc extension ('An Extension of CIDOC CRM to support social documentation') under development. This aims to:

"(re-)encode data that document social phenomena and constructs that are typically recorded by humanities and social science scholars based on their analysis and transcription of primary documentary evidence or their representation of observational data in analytic digital form" [19]

Although at an early stage of development, it has the potential to be able to record the social reception of an object/project. However, as this is not currently one of the primary areas of analysis, further work would be needed to identify if this does fall within its scope and if so any additions/changes needed to the existing proposed classes/properties in the extension to support this use.

Community

The desire to record the views of a wider audience on the musealised object/project may fall into scope for the CRMinf extension under development [20]. This aims to provide:

"a global schema for integrating metadata about argumentation and inference making in descriptive and empirical sciences such as biodiversity, geology, geography, archaeology, cultural heritage, conservation, research IT environments and research data libraries. Its primary purpose is facilitating the management, integration, mediation, interchange and access to data about reasoning by a description of the semantic relationships between the premises, conclusions and activities of reasoning."

However, the origin of this extension is primarily for academic discussion around factual statements made about the object, the desire to record a community expressing their thoughts on an object in the institutions collection [21] may be too informal for this extension and a separate extension might need to be considered if this cannot be incorporated into CRMinf.

---

[17] Amico, Nicola, and Achille Felicetti. "Ontological Entities for Planning and Describing Cultural Heritage 3D Models Creation." *arXiv preprint arXiv*:2106.07277 (2021)

[18] Developing a strategy for the conservation of performance-based artworks at Tate, Louise Lawson, Acatia Finbow, Hélia Marçal

[19] Definition of the CRMsoc, v0.1, *https://cidoc-crm.org/crmsoc/*

Related Activity

Beyond the existing extensions to CRM directly referenced (CRMdig, CRMinf, CRMsoc), there are other extensions or cataloguing developments that are interesting to consider as part of work on CRM support for born-digital cataloguing.

The work of Rhizome on cataloguing their born-digital net-art collection has led recently to the latest version of their Art Base data model [22] making use of WIkibase [23] as a linked data platform for a linked data documentation of their artworks, drawing on existing ontologies such as PROV [24] to enable artwork creation (and modifications over time) to be documented, as well as any digital preservation activity carried out.

Much work has been carried out on the usage of CIDOC CRM for documenting archaeological projects, from site documentation (CRMba) to excavations (CRMarchaeo). Of most interest for born digital, the work on documenting (pre-dominantly) archaeological cultural heritage artifacts has led to the development of proposed extensions for CIDOC CRM [17], or based on ontologies such as PROV [25].

Future Activity

The development of an interoperable standard, by its very nature, relies on the co-operation of a group of interested parties who agree on a common approach to resolve their shared needs. By the accumulation and discussion of relevant case studies which clarify these needs (and their relative importance), a standard can be developed which provides this resolution. From this survey of the four case studies from the Towards a National Collection Born Digital project, we can start to see the areas of data modelling needing further development (and collective agreement). It is to be hoped a group of interested institutions continue to accumulate and discuss further case studies, and contribute to open standards development, enabling the comprehensive born digital/hybrid cataloguing all desire.

---

[20] *https://cidoc-crm.org/crminf/*

[21] Along the lines of Revisiting Collections - *https://collectionstrust.org.uk/resource/*revisiting-collections/

[22] *https://sites.rhizome.org/artbase-re-design/data-models.html*

[23] *https://wikiba.se/*

[24] *https://www.w3.org/TR/2013/REC-prov-o-20130430/*

[25] Homburg, T., Cramer, A., Raddatz, L. *et al*. Metadata schema and ontology for capturing and processing of 3D cultural heritage objects. *Herit Sci* 9, 91 (2021). *https://doi.org/10.1186/s40494-021-00561-w*

www.vam.ac.uk