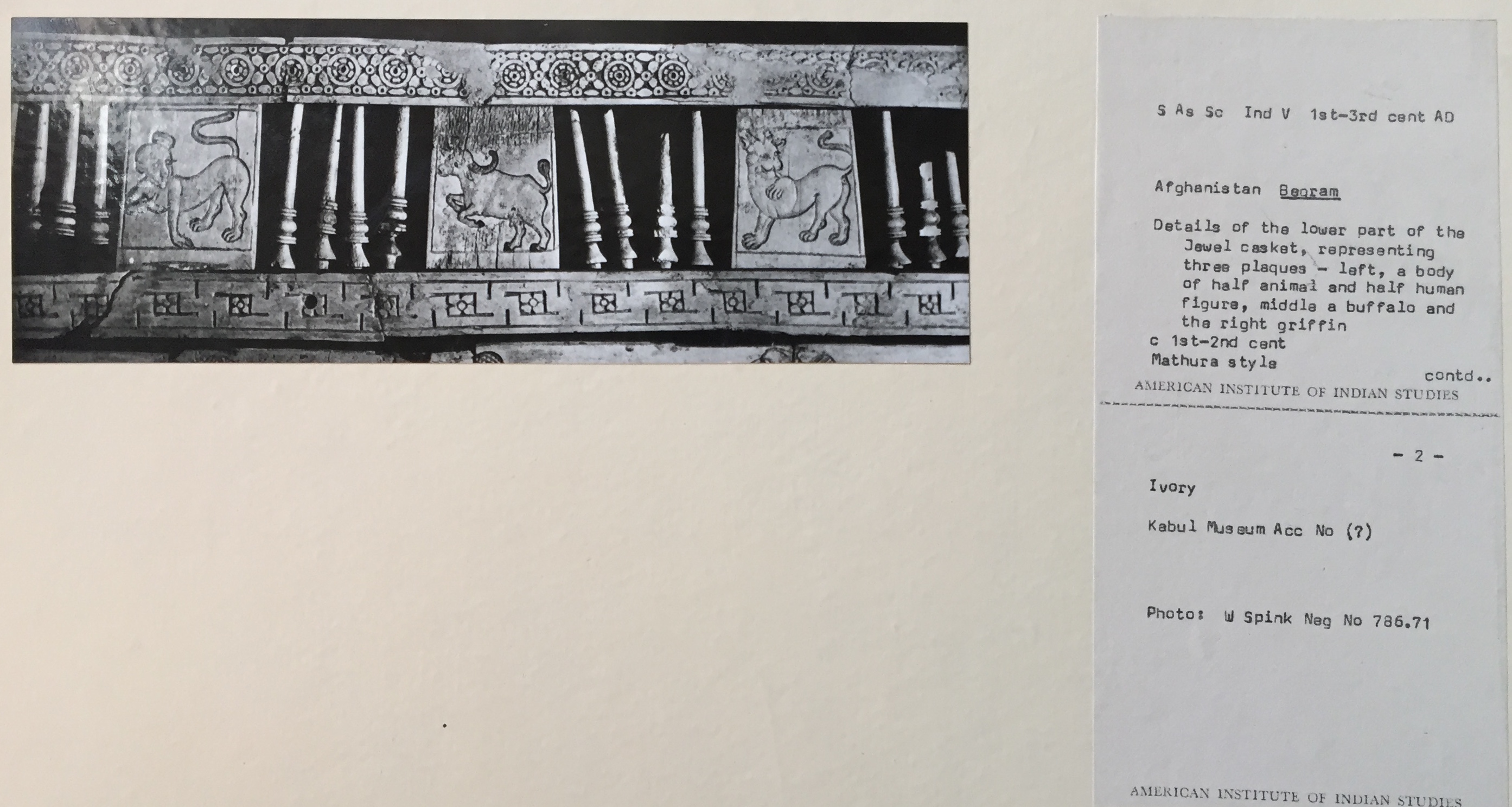
**An Explanation of the Mapping of Photographic Archives (Monument Records) of the American Institute for Indian Studies to the CIDOC Conceptual Reference Model**

Author: **Dominic Oldman and Sebastian Rahtz** with Donna Kurtz

Date: **21 April 2015**

Version: **4**





# Introduction

The *Digital Cultural Heritage*project aims to use the latest web technologies to make the highest quality research data about global heritage available to everyone.  The DCH is led Professor Donna Kurtz, Director of the *Cultural Heritage Programme* (<http://www.culturalheritage.ox.ac.uk/dch>) at the University of Oxford. In India the national coordinator is Dr Vandana Sinha in the American Institute of Indian Studies.

On 8 April 2015 AIIS hosted a Digital Cultural Heritage workshop. Sebastian Rahtz, Chief Data Architect of the University of Oxford, and Dominic Oldman, Head of the Mellon Foundation’s ResearchSpace project at the British Museum, talked about the new web technologies and heritage data. On the 9th and 10th they applied them to sample data from AIIS. They prepared this report on their initial mappings to CIDOC CRM and Linked Data Representation.

The ability to complete a CIDOC CRM (http://www.cidoc-crm.org) mapping during the short visit relied on having experts from different aspects of heritage documentation in one room - those who understood the source data schema and the information itself, and those familiar with the target model, the CIDOC CRM. The need to collaborate in this way underlines the semantic approach taken by the CRM framework and, as is the case with most source schema, pre-meeting assumptions about the data fields and their meanings proved to be inaccurate and were corrected by local experts, despite experience and familiarity with previous cultural heritage databases and ontology and technology expertise.

This report accompanies a graphical representation of the mapping to show how knowledge representation using real world references can form a semantic knowledge framework for the harmonisation of heterogeneous data sources. By “knowledge representation”, we mean the formalization of information about the world in a form that a computer program can utilize to answer questions. References to CIDOC CRM labels are in bold to make the association with this narrative more accessible for prospective modellers. After a description of each section a note is made about the harmonisation properties of the mapping.

# Process

Photographic prints in the AIIS archive are ‘traditional’ – they were taken using film cameras. The photograph is derived from a negative and given an accession number. The negative has its own identifier. However, in some cases, a digital image is also available taken using a digital camera creating a ‘born digital’ file with its own accession number; these are linked to the analogue image which is regarded as being the same view but will inevitably vary from the film version. The subsequent documentation relates to both the physical mediums on which visual information is recorded and the subject of the photograph itself, the monument. In many photographic archives this type of record is common but this division of information is often not semantically represented making it impossible for computers to differentiate between the different entities involved. The CIDOC CRM makes clear these different entities and associates contextual information correctly thereby supporting true contextual data integration.

# The Mapping Narrative

Traditional data sources are not suited to storing explicit contextual information and data is reduced to related lists of fields often resulting in flat linear data representation. The mapping described in the document transforms a spreadsheet table into a hierarchical framework that restores its natural or ‘real world’ meaning. The labels used in the mapping are fully and scientifically described in the CIDOC CRM reference manual and their use is governed by strict semantic rules. The following key entities are included in the AIIS mapping.

**Information Carrier** – Information is held on many different physical mediums but the information that they hold can relate to the same or related things. The general entity Information Carrier allows alignment of the different physical things that may contain related information, for example, a photograph, a photographic negative, a book or a computer disk. The physical carrier can depicts a visual item (a picture) and/or carry other symbols like text.

**Information Object (Document) –** The actual information contained on the physical media is intellectually distinct from it physical carrier and as an immaterial item it has different properties. Information can be created through text, images or even memories. The type of information object in this case is a ‘document’ (in the cultural heritage sense) and documents a physical thing.

**Production** – A key aspect of the CIDOC CRM is that it is event based. It works by recording the interaction of things with a permanent identity (whether material or immaterial) with events in time distinguished by their temporal nature. A production event allows us to include dating information, or information about how and where the thing was produced.

**Type** – All documentation systems use terminologies or concepts. The CRM allows these concepts to be included in their original form without change or homogenization thereby reflecting local descriptions and language. Examples in the mapping include terms for the materials used in construction, the type of architecture or thing, the subjects assigned by the AIIS, and so on.

**Place** – The geographical location may include different appellations. Places are often hierarchical in nature and the relationship “falls within” allows the relationships between places to be recorded. Multiple appellations can record place names and coordinates.

## Photographs and Negatives

The physical media, the negative and the photographic print, are **Information Carriers** (they are physical media that carry an **Information Object** (the information derived from the image). In this case there are initially two types of **Information Object,** a **Document** and a **Visual Image.** Thephotographic **Information Carrier** carries a **Document** in that it ‘documents’ (in the cultural heritage sense) the monument but it also **shows** a **visual item** which is the **Visual Image** of the monument. The **Document** is a record of information about the monument which itself is a **Physical Man-made Thing**. The **Visual Image** more precisely **‘represents**’ the monument. The negative and the photograph carry and represent the same **Information Object** and **Visual Image. Both** the **Information Carriers** (the photograph and the negative), as physical items, **depict** the **Physical Man-made Thing.**

Where a digital image is also captured this also creates a **Document**, in this case a **Digital Object** (another type of **Information Object**) of the monument. The **Digital Object** is also stored on a physical medium, an **Information Carrier** (a disk) which also shows a **Visual Item** and results in the **documentation** and **representation** of the same **Physical Man-made Thing**.

|  |
| --- |
| Data Harmonisation Notes These important relationships promote multiple routes for integrations with other types of documentation of the same **Physical Man-made Thing** (text, video, etc.) but also make a more specific integration with other items that show a visual representation including other photographs but also other information carriers whether they be 3D or 2D objects (paintings for example) with illustrations of the same subject. Integration with things that are connected in a more conceptual way would also be supported. |

## The Photo Production Process

The archive records provide the date that the photograph (**Information Carrier**) was produced. The photograph is produced by **using a specific** negative (**Information Carrier**). The **Production** of this is recorded within a **Time-span** with a date range. Within this process the photographic print is given an accession number, an **Identifier**). Equally the negative has a negative number, another form of **Identifier.** The **Digital Object** also has an **Identifier**.

|  |
| --- |
| Data Harmonisation Notes These unique identifiers are exposed as Linked Data and allow other data sources to make reference to them supporting contextual integration. |

## Additional Photographic Information

The remaining photographic information records the owner (the **Actor**: AIIS), the category or **Type** of the photograph (e.g. architecture), whether the photograph was taken from the exterior or the interior (another **Type** or terminology), and information used to describe the photograph, its label information, which is another form of **Document**.

|  |
| --- |
| Data Harmonisation Notes This information is important additional context that facilitates harmonisation. There is no guarantee that the description given, for example, the name of the monument, uses the same appellation or that the name is spelt the same way. This additional information provides a way of mutually corroborating the record and can support the main contextual framework to match instances from different databases that may contain different appellations or spellings. |

## The Physical Thing

Once we have established the subject of the photograph, either through **documentation**, **representation** or **depiction,** the record then deals with the **Physical Thing** itself. It has its own set of terminological categories. The monument **Type** describes the purpose of the monument such as a palace or a tomb and so on, while another **Type** provides the style of architecture. The **Materials** that the building consists of is another more specific **Type.** The record also describes the **Period** that the monument dates from (the period of its **Production**) and the patron of its **Production** (The **Person** or **Group** who performed that role) - the **Actor.** The approximate time scale for the **Production** of the monument is also recorded using a **Time-span** and date range, which should be within the described **Period.**

|  |
| --- |
| Data Harmonisation Notes The **Production** of material things provides important references for integration, not only with data about the same thing but also with other types of thing that have some relevance or provide some additional context. For example, objects that might have been produced in the same period and which might have had some connection such as items that might have been found with or within the thing in question. This provides historical context that helps improve the relevance of the record *and* further context for additional data harmonisation with other data sources of different types, for example, reference materials. Connections with people are also important. They can provide a additional information useful for research and integration, particularly if connections (perhaps using the material context) can be made with specialist biographical resources. |

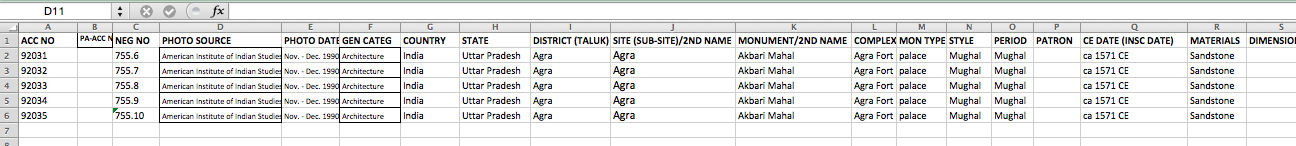
## Location or Place

Location information is recorded as a set of hierarchical places each of which **falls within** the geographical area of the last. The monument as a place, ‘**falls** **within’** the complex, the complex  **‘falls within’** the site, the site ‘**falls** **within’** the District, the District ‘**falls within’** the State, and the State ‘**falls within’** the Country. **Spatial coordinates** are separately used to identify the key place name. Each place is associated with a particular appellation for that place, and additional alternative appellations (with **Time-spans**) can be added.

|  |
| --- |
| Data Harmonisation Notes Each level of location information provides a potential reference point for further integration with or without further data. Different records will provide different levels of location information but also may provide different place appellations. Places can be known by different names and different names can be recorded in the same record. The location information, together with other contextual information, can be used to support matching with other sources that use different place names, building up a thorough knowledge of the different appellations that have been allocated over time, both archaic and modern. |

# Appendix: Source data and conversion

The archive records are maintained using Excel spreadsheets, with column headings indicating the data categories. An example is shown below; note that the data (e.g., ‘Country’) are repeated for each entry, leaving open the possibility of inconsistent data entry.



Using a system developed at Oxford for document conversion (OxGarage: <http://oxgarage.oucs.ox.ac.uk:8080/ege-webclient/>), we were able to convert this spreadsheet to an intermediate XML form, which we then transformed to a flat XML structure in which each of the column headings is used to make an XML element, as in the following example:

<record>

<acc\_no>92032</acc\_no>

<neg\_no>755.7</neg\_no>

<photo\_source>American Institute of Indian Studies</photo\_source>

<photo\_date>Nov. - Dec. 1990</photo\_date>

<gen\_categ>Architecture</gen\_categ>

<country>India</country>

<state>Uttar Pradesh</state>

<district\_taluk>Agra</district\_taluk>

<site\_sub-site2nd\_name>Agra</site\_sub-site2nd\_name>

<monument2nd\_name>Akbari Mahal</monument2nd\_name>

<complex>Agra Fort</complex>

<mon\_type>palace</mon\_type>

<style>Mughal</style>

<gen\_categ>Architecture</gen\_categ>

<country>India</country>

<state>Uttar Pradesh</state>

<district\_taluk>Agra</district\_taluk>

<site\_sub-site2nd\_name>Agra</site\_sub-site2nd\_name>

<monument2nd\_name>Akbari Mahal</monument2nd\_name>

<complex>Agra Fort</complex>

<mon\_type>palace</mon\_type>

<style>Mughal</style>

<period>Mughal</period>

<ce\_date\_insc\_date>ca 1571 CE</ce\_date\_insc\_date>

<materials>Sandstone</materials>

<exterior\_\_interior>Exterior</exterior\_\_interior>

<subject\_label>General view of the ruined structure</subject\_label>

<subjectsearchkeyword>Akbari mahal</subjectsearchkeyword>

<subjectsearchrelatedkeyword>Akbar's palace, Akbar palace</subjectsearchrelatedkeyword>

<geocoordinates>27° 5′ 0″ N, 77° 58′ 0″ E</geocoordinates>

<nearby\_bus\_stand>Agra</nearby\_bus\_stand>

<nearby\_railway\_station>Agra </nearby\_railway\_station>

<nearest\_airport>New Delhi</nearest\_airport>

</record>

A conversion of this format to RDF format using the CRM ontology is now relatively straightforward, and results in triples like the example below, which shows some of the key entities and relationships for the accession number 92032. It should be clear from that this that there is a considerable issue to be resolved over which URLs to be used for AIIS entities.

@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .

<http://www.indiastudies.org/AIIS/photo/92032>

a crm:E83\_Information\_Carrier ;

crm:P108i\_was\_produced\_by <http://www.indiastudies.org/AIIS/production/novdec1990> ;

crm:P128\_carries <http://www.indiastudies.org/AIIS/keyword/akbarpalace> , <http://www.indiastudies.org/AIIS/keyword/akbarspalace> ;

crm:P138i\_has\_representation <http://dsal.uchicago.edu/images/aiis/images/large/ar\_092032.jpg> ;

crm:P2\_has\_type <http://www.indiastudies.org/AIIS/category/architecture> , <http://www.indiastudies.org/AIIS/intext/exterior> ;

crm:P52\_has\_current\_owner <http://www.indiastudies.org/AIIS/actor/americaninstituteofindianstudies> ;

crm:P62\_depicts <http://www.indiastudies.org/AIIS/place/akbarimahal> ;

crm:P70i\_is\_documented\_in <http://www.indiastudies.org/AIIS/photolabel/92032> .

<http://www.indiastudies.org/AIIS/photolabel/92032>

a crm:E31\_Document ;

rdfs:label "General view of the ruined structure" .

<http://www.indiastudies.org/AIIS/subject/akbarspalace>

a crm:E55\_type ;

rdfs:label "Akbar's palace" .

## Appendix - Graphical Representation

