UK Research and Innovation

Arts and Humanities Research Council

British Library

Royal Botanic Garden Edinburgh

The National Gallery

University of Glasgow

V&A

Science Museum Group

Natural History Museum

Towards a National Collection: Persistent Identifiers as IRO Infrastructure

Project Launch Webinar, Edinburgh: 6th April 2020

Joseph Padfield (NG)

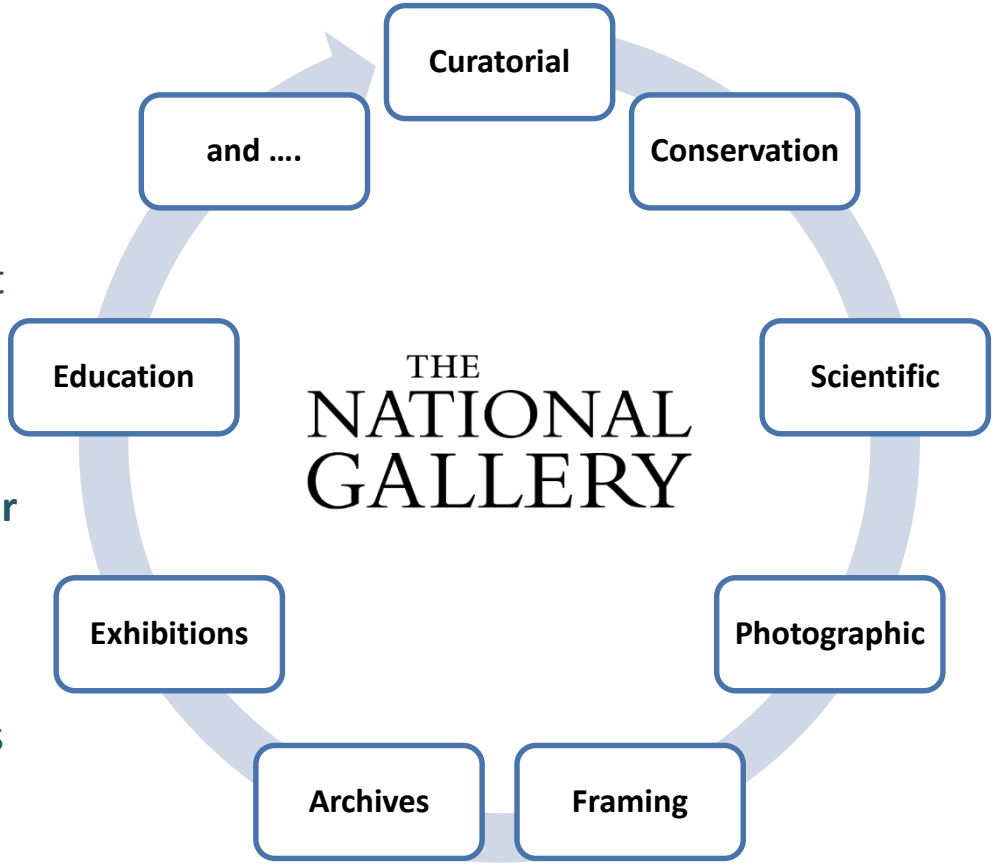
Working towards Persistent Identifiers in the National Gallery, London.

- Over 2,300 main collection works of art in the European Style.
- Ranging from the 13th to the early 20th century.
- Arranged over 66 public Galleries.
- Including many famous works, such as
  - van Eyck's Arnolfini Portrait
  - Velázquez's Rokeby Venus
  - Turner's Fighting Temeraire
  - Van Gogh's Sunflowers.

National Gallery

National Gallery – Research Resources

- **Wide range of resources** of National Gallery research content – everything from bibliographies to images and spectra.
- No single system, but a range of different **systems, resources, people and knowledge**.
- An early question: "**what could be used to connect all of this knowledge together and then potentially connect it to external resources?"**
- **For a number of years** we have been exploring how we might connect systems together and improve access.



UK Research and Innovation

Arts and Humanities Research Council

British Library

Royal Botanic Garden Edinburgh

The National Gallery

University of Glasgow

V&A

Science Museum Group

Natural History Museum

Physical Access to NG Archives has been common for many years

ARCHLAB: Physical Access to the Scientific Archive

Archives comprising thousands of unpublished images, data, samples and reports collected over many decades of scientific investigation and conservation of a huge variety of cultural heritage objects and sites

AND

The specialised knowledge at the staff of the ARCHLAB institution to guide the researcher through the archive

ARCHLAB provides EU funded opportunities to explore scientific archives.

ARCHLAB (NG) - *The Capodimonte Transfiguration: a technical and historical study of a masterpiece by Giovanni Bellini in context*

Angela Cerasuolo , Alessandra Rullo, Museo e Real Bosco di Capodimonte; Helen Glanville, LAMS – CNRS Sorbonne Universités

IPERION CH

2009 CHARISMA 2015 IPERION CH 2020 IPERION HS 2022 E-RIHS EUROPEAN RESEARCH INFRASTRUCTURE FOR HERITAGE SCIENCE

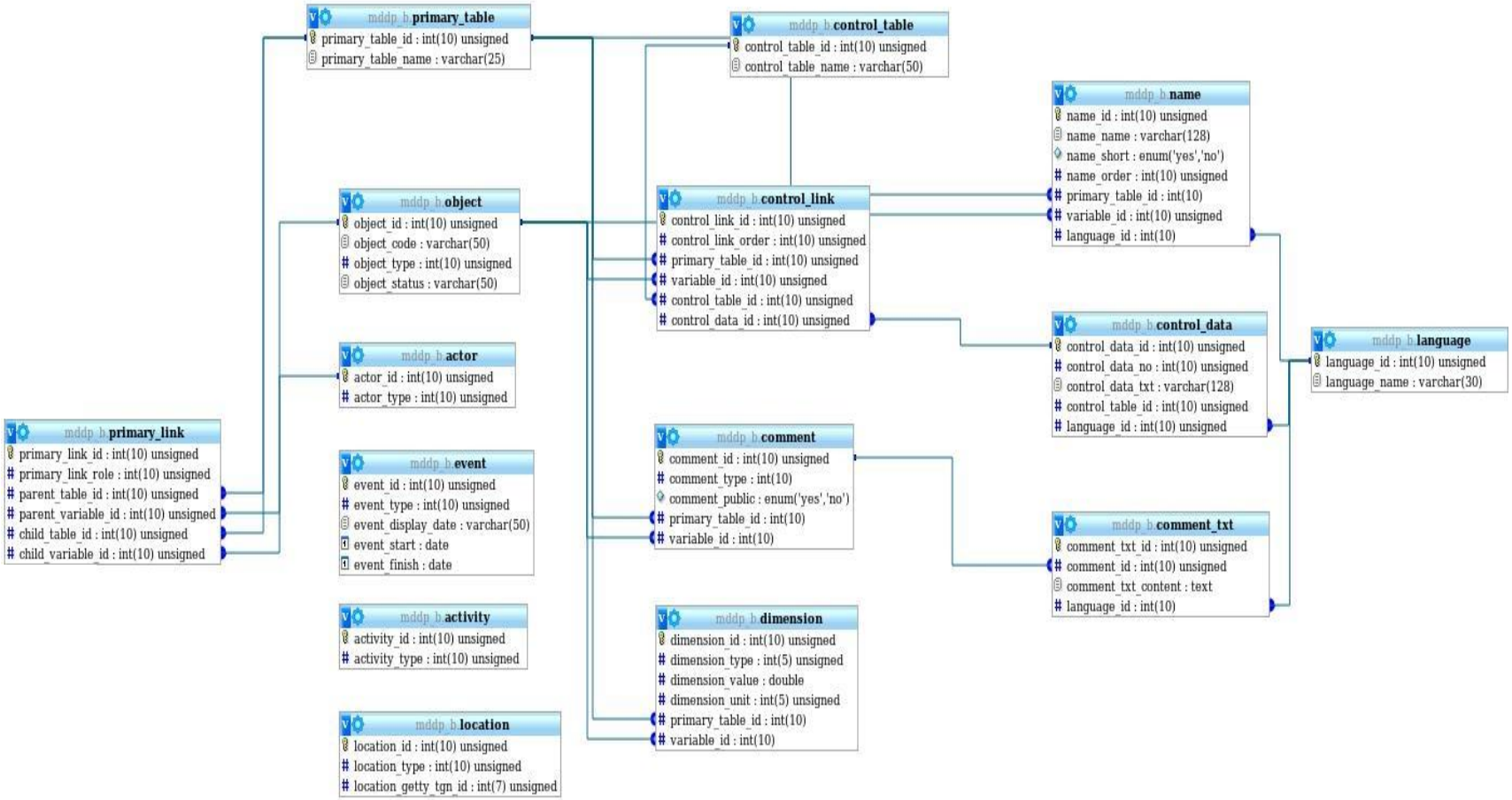
UK Research and Innovation | Arts and Humanities Research Council

British Library | Royal Botanic Garden Edinburgh | The National Gallery | University of Glasgow | V&A | Science Museum Group | Natural History Museum

Raphael Project 2007: Database Development

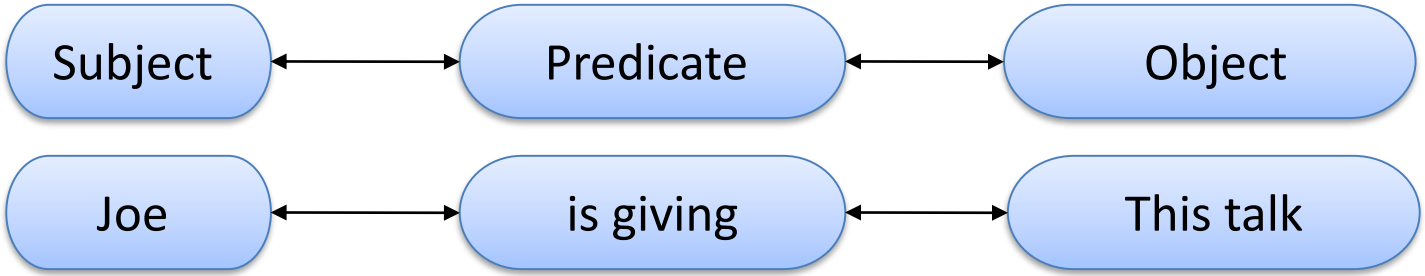
- For this prototype system a MySQL database was developed to store the wide range of data.
- Initial development produced a generic 15+ table schema.

Raphael Research Resource: Initial Scheme



Semantically describing paintings using RDF:

- After further discussion this simplification process led to an alternative way or storing data.
- The RDF data model is based upon the idea of making statements about resources in the form of subject-predicate-object expressions, called triples in RDF terminology.
- The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object.
- Ideally, each subject, predicate and object require a *unique identifier* that can be *re-used and referenced*.

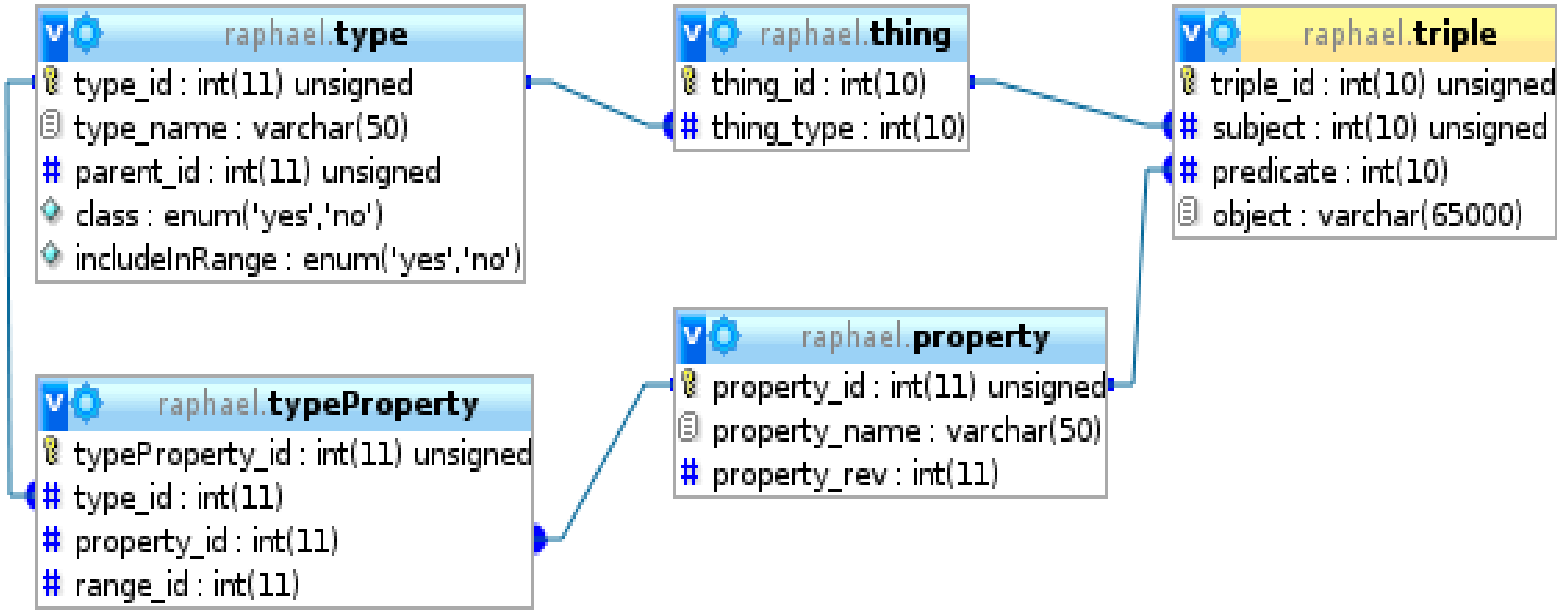


Raphael Project 2007: Database Development

- For this prototype system a MySQL database was developed to store the wide range of data.
- Initial development produced a generic 15+ table schema.
- Experimentation with various Resource Description Framework (RDF) based systems allowed the schema to be simplified even more to 5-6 tables.
- At the time, full migration to an RDF-triple based system was not possible due to time constraints and integration issues with some of the tools used in the graphical user interface.

Raphael Research Resource: Database development

- Within this pilot project it was decided to make use of this different approach to storing data.



Raphael Research Resource

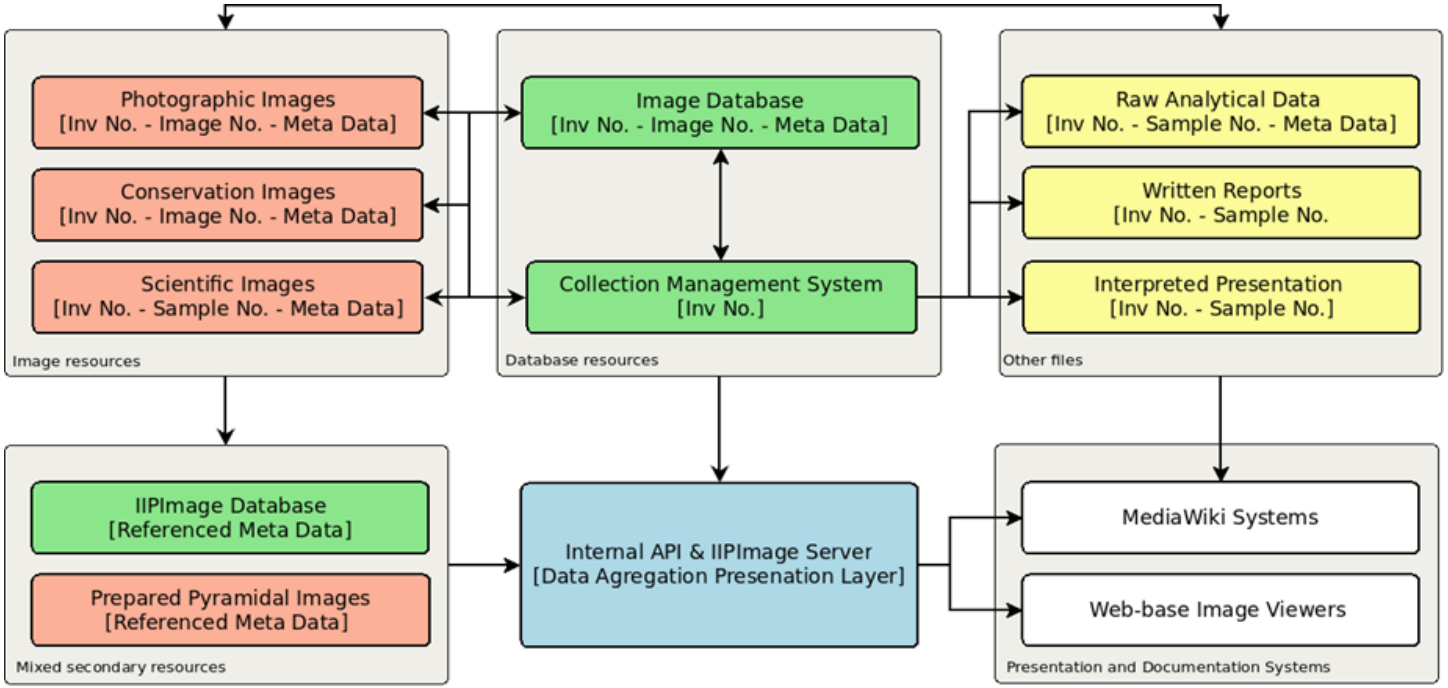
Initially the project was based on the extensive studies of the ten paintings by Raphael in the Gallery collection, but it was extended to hold information provided by a several other institutions and collaborators.

- But the system was designed as a graphical user interface for humans to use.
- It was not possible to reference individual paintings let alone individual data points.
- The system made use of its own internal identifiers which where not designed for re-use.

http://cima.ng-london.org.uk/documentation

National Gallery – Connecting Data for Reuse

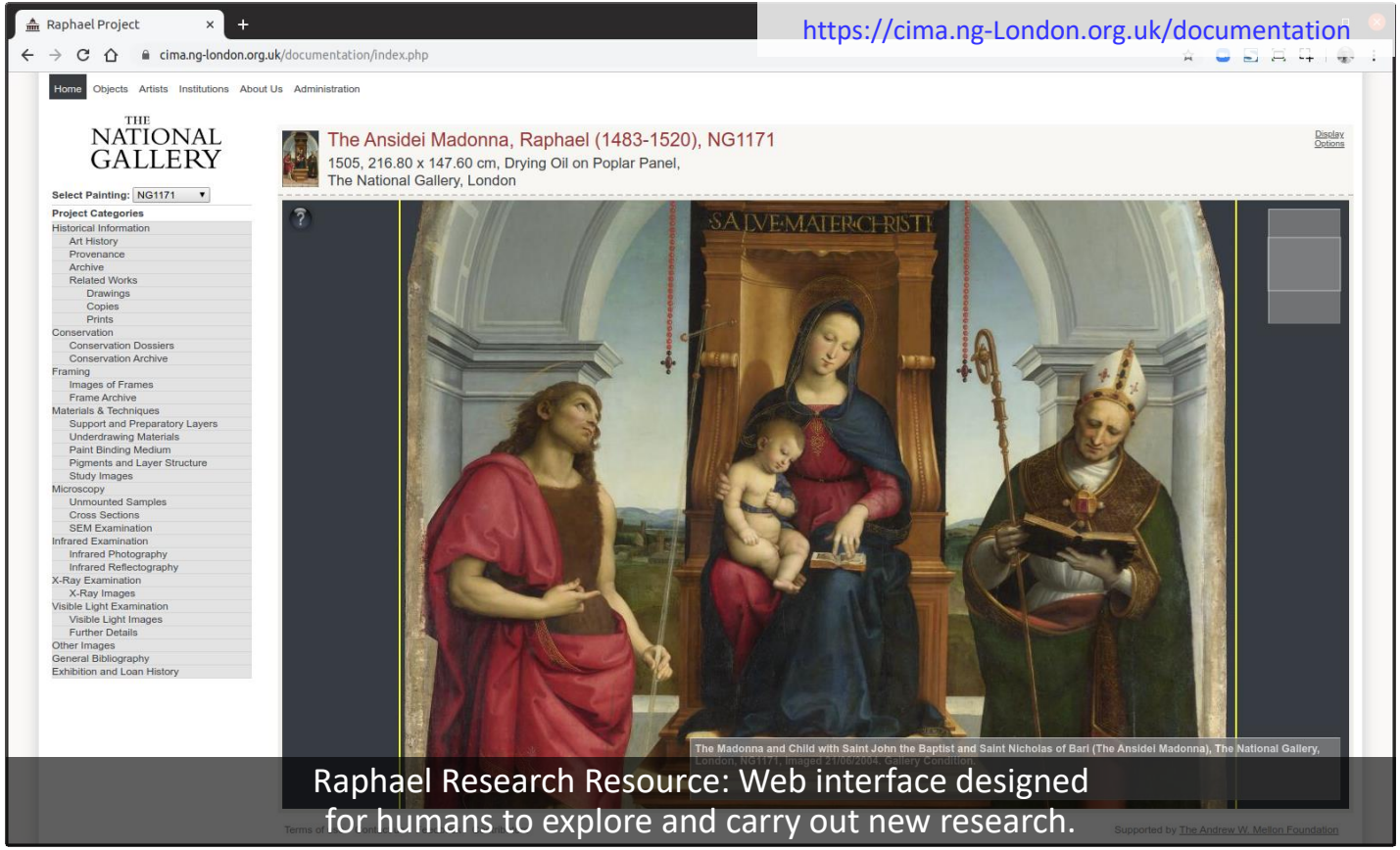
Over the same period internal naming conventions allowed a number of resources to be identified, classified and automatically linked, via: Painting Inv. No. - Image or Sample No. & Technique



Heritage Digital Identifiers

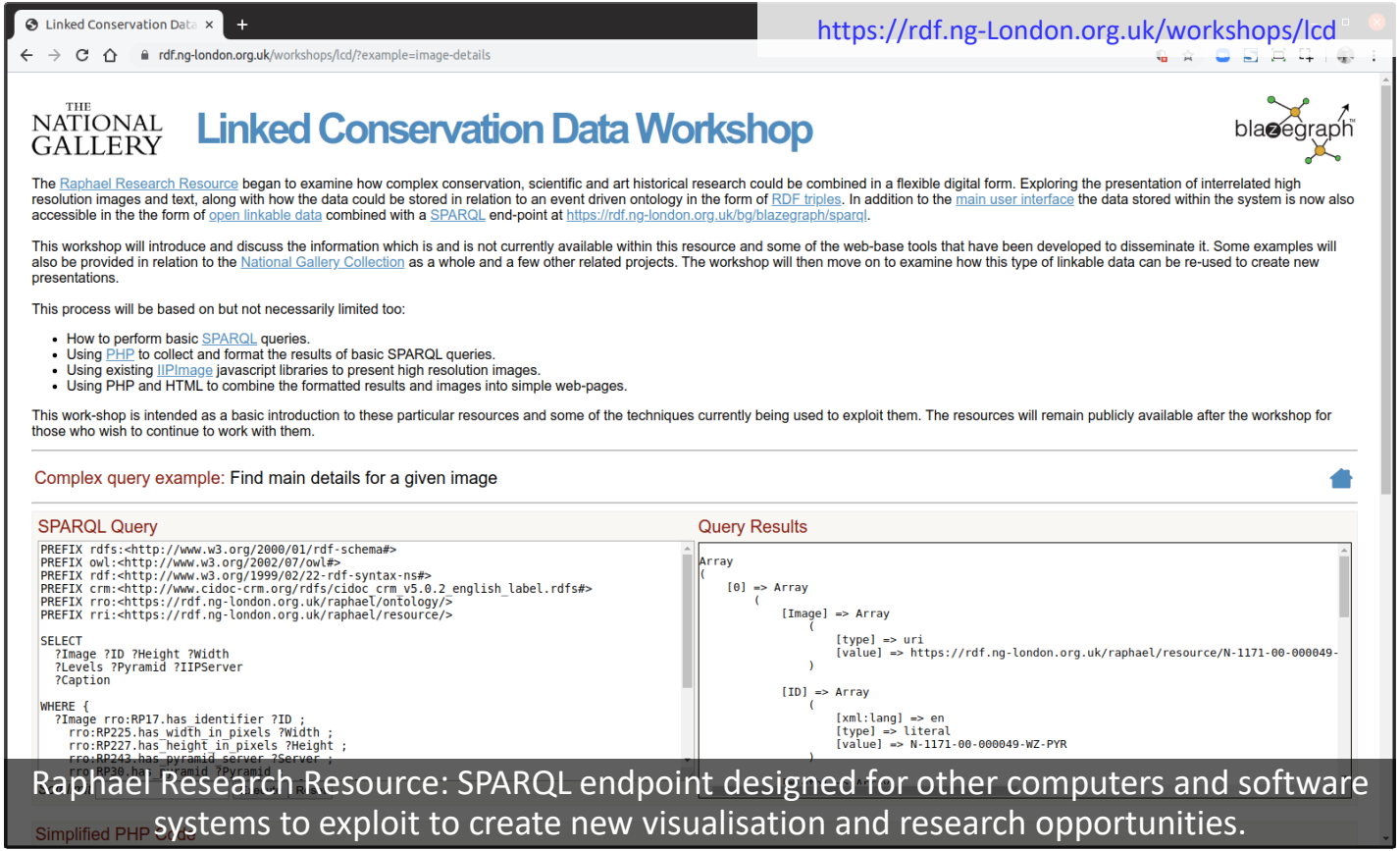
- NG Paintings are organised by Inv. No. However this number relates to a paintings state and what part of the collection a painting is in, so it is subject to change.
    - Most digital images are directly named in relation the current Inv. No. – so these are also subject to change.
    - Other image identifiers, often include semantic state information and even meta data - there are also different rules and conventions based on the type and source of an image.
- Other database IDs are directly linked to specific pieces of software – so their use, between systems, can involve a risk if software needs to change.
- More broadly, accession numbers, in museums are also commonly based on the year of acquisition, therefore the use of these simple IDs across multiple institutions becomes difficult.

Designed For Human Re-use



Raphael Research Resource: Web interface designed for humans to explore and carry out new research.

Later work: Re-designed For Computer Re-use



Raphael Research Resource: SPARQL endpoint designed for other computers and software systems to exploit to create new visualisation and research opportunities.

Development of the SPARQL end-point highlighted that further work was required to improve the usefulness of this type of structure and begin to consider how it might be used to present the full collection.

UK Research and Innovation

Arts and Humanities Research Council

British Library

Royal Botanic Garden Edinburgh

The National Gallery

University of Glasgow

V&A

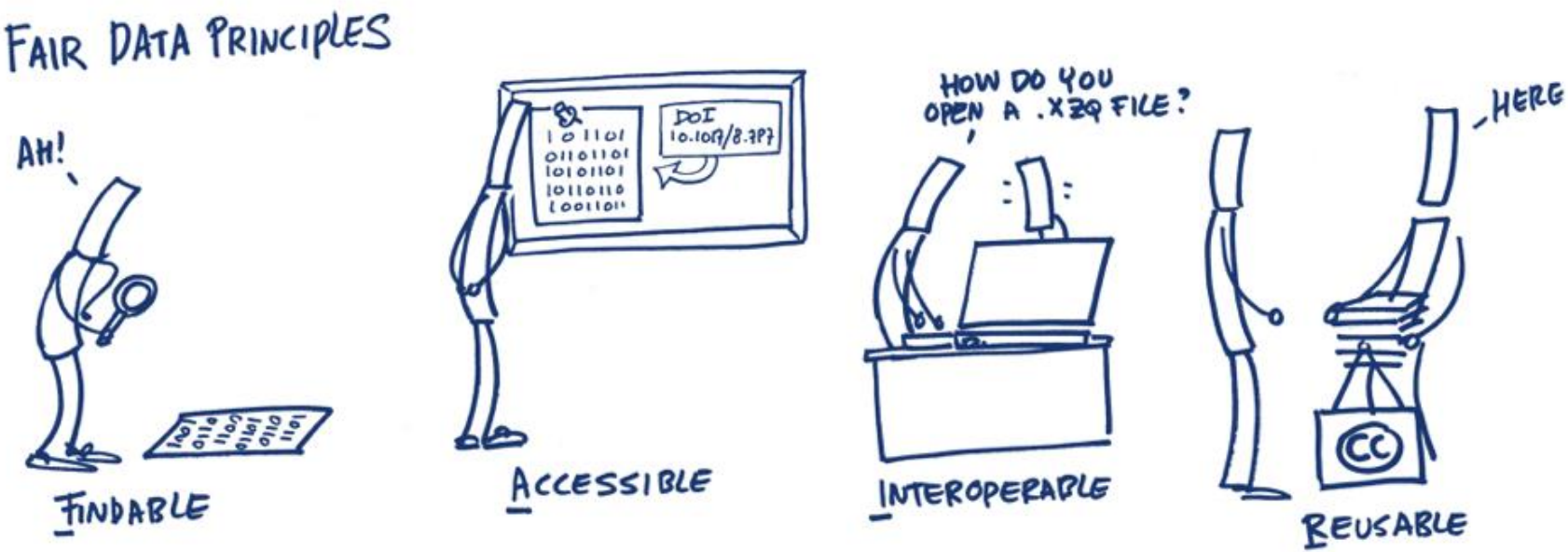
Science Museum Group

Natural History Museum

But beyond this one off project how could we improve access to digital data?

The Standardising Sharing: F.A.I.R. Data Principles

https://www.go-fair.org/fair-principles/



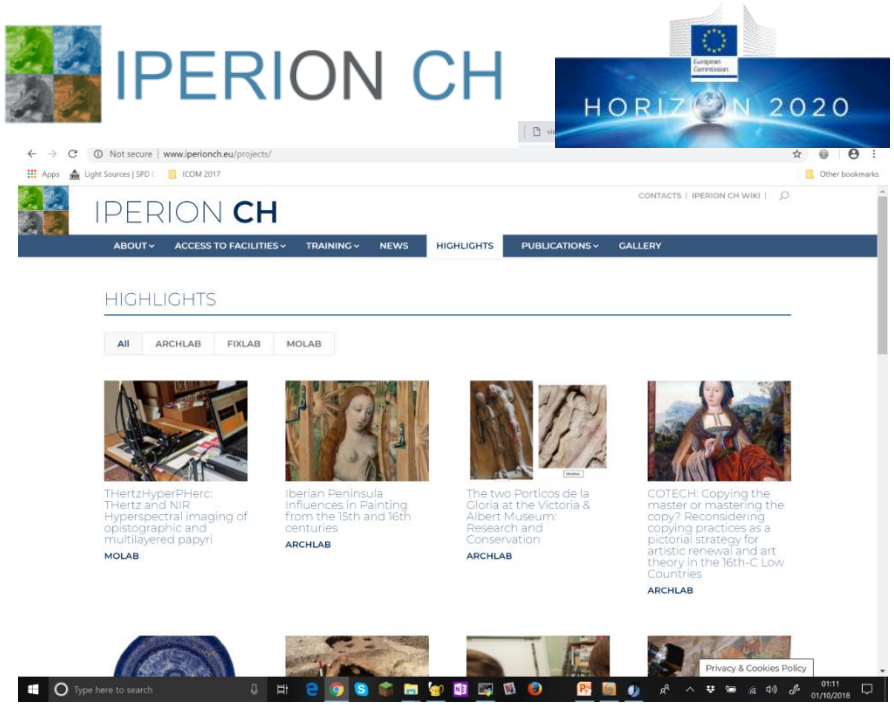
https://www.openaire.eu/images/Guides/FAIRdataprinciples_foster.png

Need:

- Simple PID registry
- Identify available data – painting and artist details, images, etc.
- Build and document an open API, with resolvable URIs, using the CIDOC CRM to map data and IIIF to present images

What was required to make our data accessible

Integrated Platform for the European Research Infrastructure ON Cultural Heritage



http://www.iperionch.eu/

- 4 year project, with 24 partners and more than 35 research groups from 12 EU Member States and the USA
- 3 Joint Research Work packages
- 3 Core Accesses: ArchLab, FixLab & MoLab

- One task was specifically exploring developments in digital documentation

National Gallery: Beta PID System.

- Basic URL Format - NG API
- NG data URLs take the form:

- **namespace/AAA-AAAA-AAAA**
- where:
- The namespace is:
  - https://data.ng-london.org.uk/resource/
- A is any alphanumeric character and together they represent a unique NG PID.

- The remaining characters are arbitrary and just form the rest of the PID.
  - Initially we will be using 8 additional characters: which will allow for 36^8 possibilities per prefix.
  - The system has been built to allow additional sets of 4 characters to be added to scale the number further if required.

In this initial development stage the first set of three characters act as a prefix, indicating what kind of entity is being described by the PID.

National Gallery: Beta PID System.

- Automated process for incrementing the last used PID using digits 0-9 and upper-case letters A-Z.
- The four character groups are considered from left to right but are incremented right to left.
- If the final 4 character group increments back round to "0000" a new 4 character group will be added on to the right, beginning at "0001".

For example(s):

001-0001-0000 => 001-0002-0000

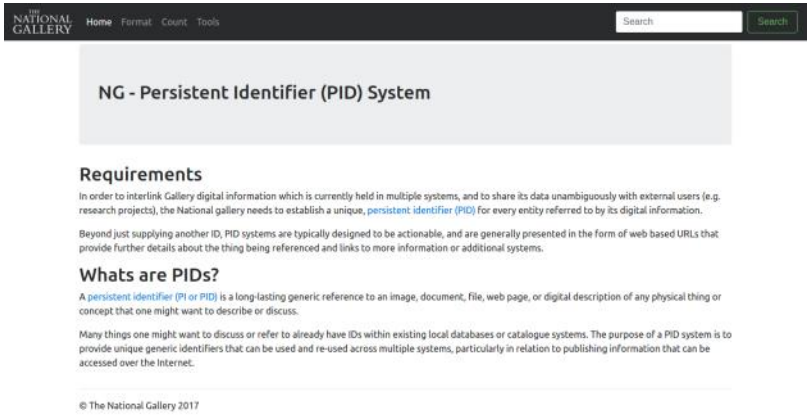
001-000Z-0000 => 001-0010-0000

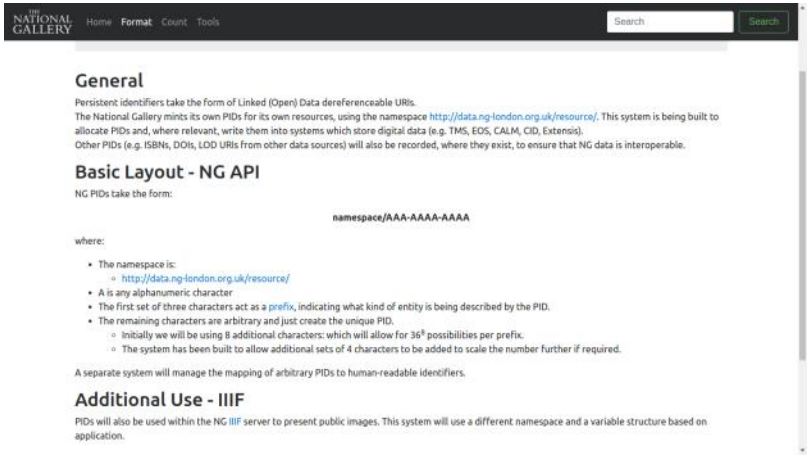
001-0019-0000 => 001-001A-0000

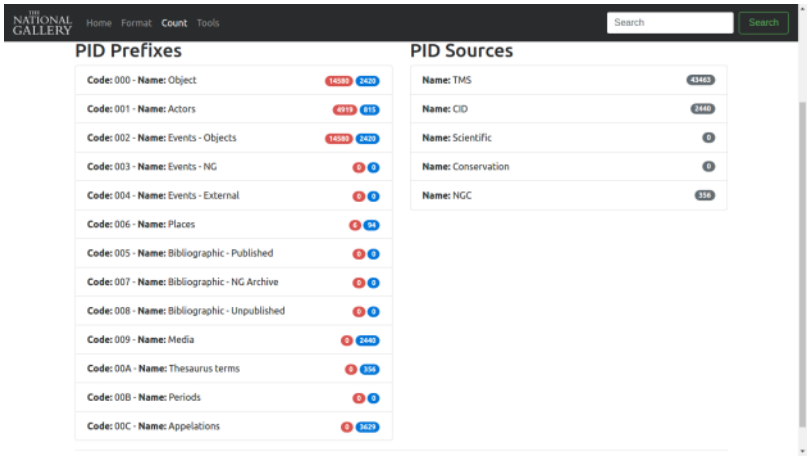
001-ZZZZ-00A1 => 001-0000-00A2

001-ZZZZ-ZZZZ => 001-0000-0000-0001

National Gallery: Beta PID System.







PIDs generated and then fed back into NG systems, such as TMS.

- NG1 - 000-03JR-0000
- Duccio - 001-0145-0000
- Room 51 - 006-000O-0000

API - Json Examples

Introduction

The Json formatted data presented by this Beta API is currently under development, this page presents some examples of how the data will be formatted.

The Json formatted data is relativly stable but still might be subject to changes and should only be used for testing purposes.

Object Details: 000-03JR-0000

{"type":"object","pid":"000-03JR-0000","no":"NG1","pdate":"1517-19","artist":{"display_name":"Sebastiano del Piombo incorporating designs by Michelangelo","artist_details":{"Sebastiano del Piombo":{"ArtistDisplayDate":"about 1485 - 1547","birth":{"display":"About 1485","start":"1483-01-01","end":"1487-12-31"},"death":{"display":"1547","start":"1547-01-01","end":"1547-12-31"}}}},"group":"","ftitle":"The Raising of Lazarus","stitle":"The Raising of Lazarus","materials":"Oil on canvas, transferred from wood","medium":["oil"],"support":["canvas","wood"],"height":"381.000","width":"289.600","xp":0,"yp":0,"credit":"Bought, 1824","location":{"pid":"006-000C-0000","name":"Room 8"},"wtext":"
The subject of this painting is taken from the New Testament (John: 11). At the request of the sisters Martha and Mary, Jesus visits the grave of their brother Lazarus and raises him from the dead.

This work was painted for Cardinal Giulio de' Medici in Rome in competition with Raphael's 'Transfiguration', now in the Vatican Gallery. It was subsequently taken to the Cathedral of Narbonne. Some of the main figures are based on drawings which Michelangelo supplied for the guidance of Sebastiano.

National Gallery - Beta IIIF Sever

mirador

Purpose

Welcome to the NG Beta International Image Interoprability Framework (IIIF) Server, which has been set-up to allow direct access to a limited set of National Gallery images.

These web pages have been set up to provide examples of how the system works in a human viewable form, displaying images in the Project Mirador viewer, but the system is generally designed to be called directly by other computer systems.

A simple search engine has been setup to search for images based on PIDs, names, title, inventory numbers, rooms, keywords, etc.

Search Search

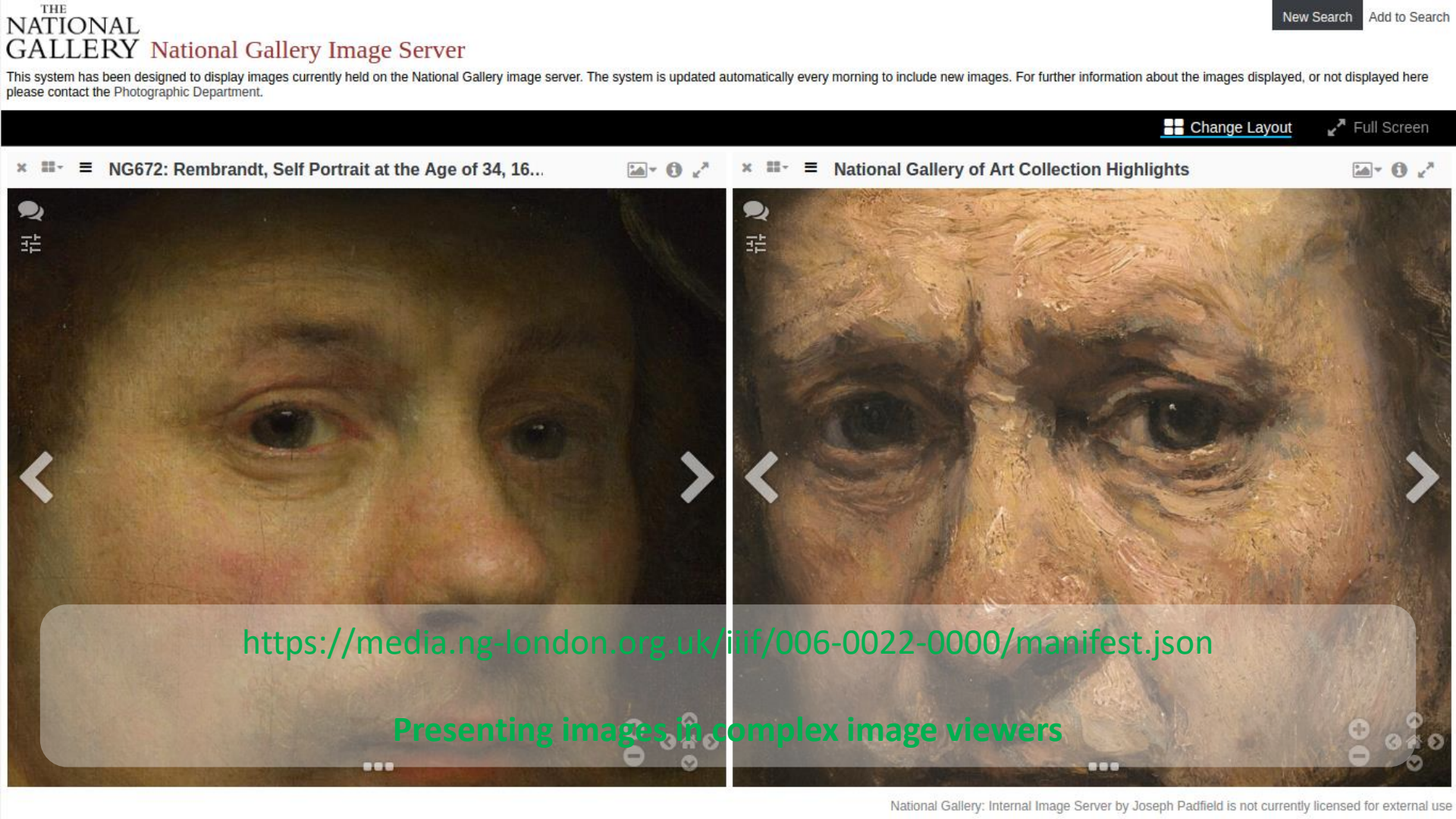
Basic URL Format - NG IIIF Server

NG IIIF URLs take the form: **namespace/AAA-AAAA-AAAA/iiif-variables**, where:

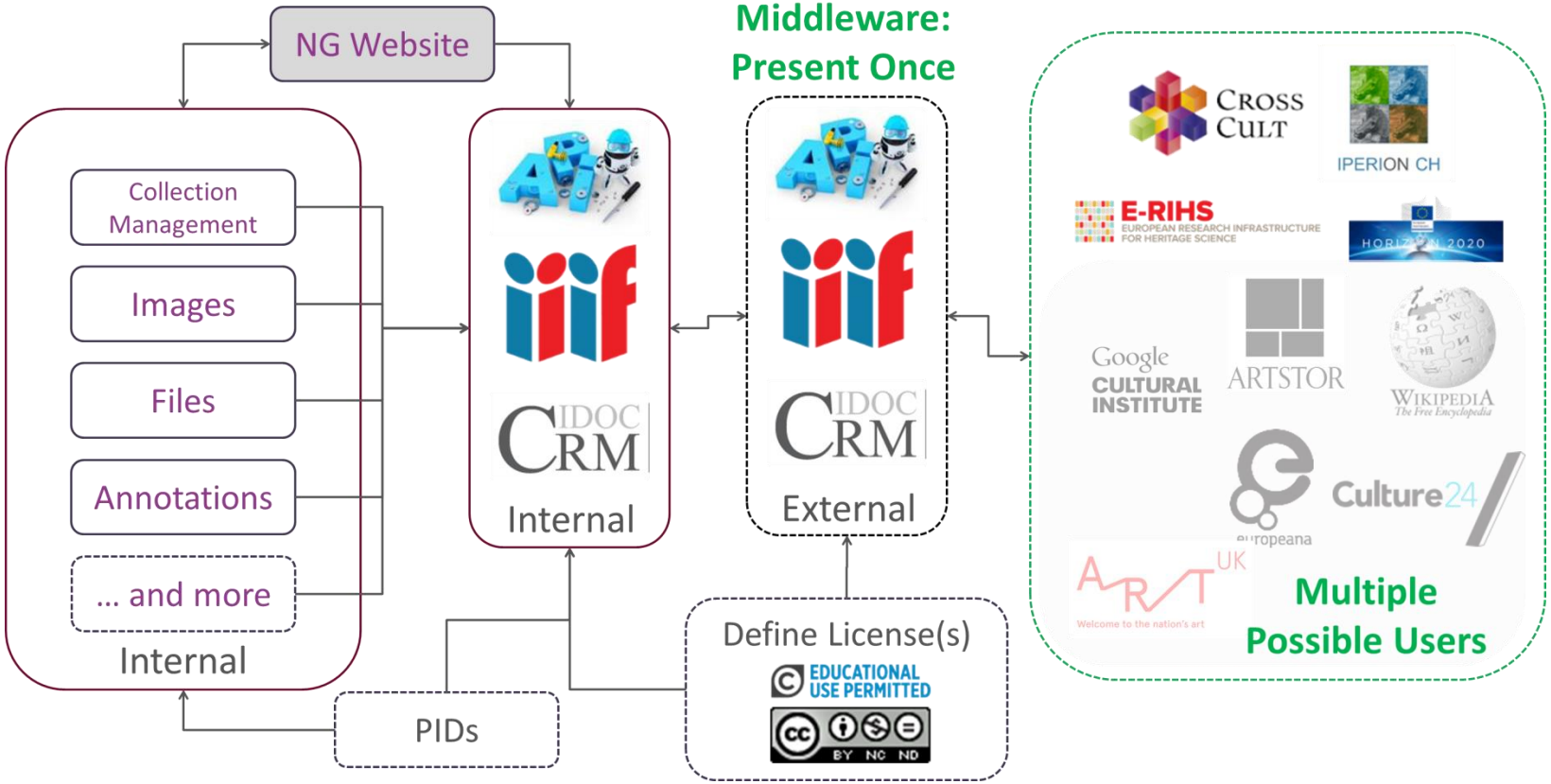
- The namespace is:
  - http://media.ng-london.org.uk/iiif/
- A is any alphanumeric character and together they represent a unique NG PIDs.

http://media.ng-london.org.uk/iiif/009-0039-0000/full/512,/0/default.jpg

Accessing simple jpegs


THE NATIONAL GALLERY
National Gallery Image Server
New Search
Add to Search
This system has been designed to display images currently held on the National Gallery image server. The system is updated automatically every morning to include new images. For further information about the images displayed, or not displayed here please contact the Photographic Department.
Change Layout
Full Screen
NG672: Rembrandt, Self Portrait at the Age of 34, 16...
National Gallery of Art Collection Highlights
https://media.ng-london.org.uk/iiif/006-0022-0000/manifest.json
Presenting images in complex image viewers
National Gallery: Internal Image Server by Joseph Padfield is not currently licensed for external use

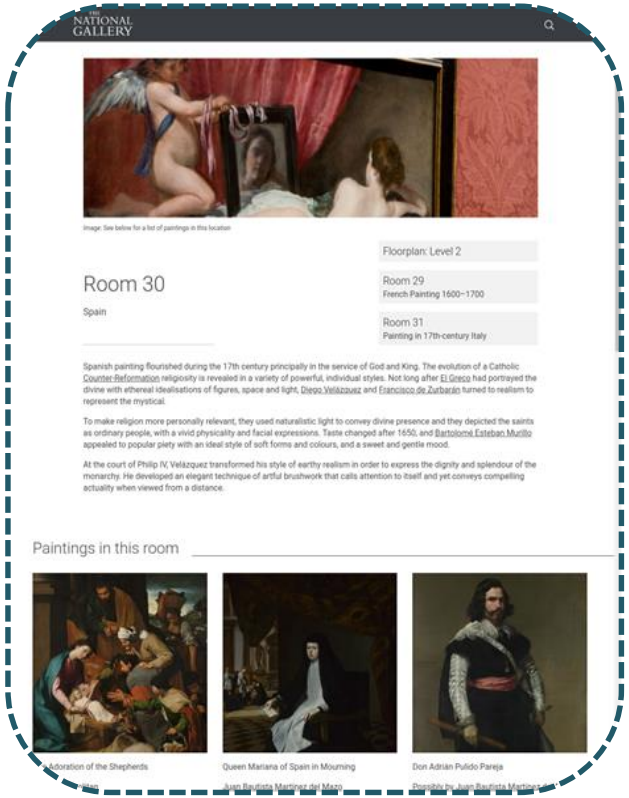
National Gallery – Aggregating internal data for re-use



Cross Cult

https://data.ng-london.org.uk/resource

Multiple standards based presentations of the same data



Formatted Website


```
object {13}
  type : location
  pid : 006-001M-0000
  name : Room 30
  title : Spain
  description : <p>Spanish painting flourished during the 17th
                century principally in the service of God and
                King. The evolution of a Catholic Counter-
                Reformation religiosity is revealed in a variety
                of powerful, individual styles. Not long after El
                Greco had portrayed the divine with ethereal
                idealisations of figures, space and light, Diego
                Velázquez and Francisco de Zurbarán turned to
                realism to represent the mystical.</p>\n<p>To make
                religion more personally relevant, they used
                naturalistic light to convey divine presence and
                they depicted the saints as ordinary people, with
                a vivid physicality and facial expressions. Taste
                changed after 1650, and Bartolomé Esteban Murillo
                appealed to popular piety with an ideal style of
                soft forms and colours, and a sweet and gentle
                mood.</p>\n<p>At the court of Philip IV, Velázquez
                transformed his style of earthy realism in order
                to express the dignity and splendour of the
                monarchy. He developed an elegant technique of
                artful brushwork that calls attention to itself
                and yet conveys compelling actuality when viewed
                from a distance.</p>
▶ objects {21}
  example_object : 000-01D6-0000
▶ artists {3}
▶ date_range {2}
▶ contains [0]
▶ keywords {63}
  license : https://creativecommons.org/licenses/by-nc-nd/4.0/
  attribution : This data is licensed under a Creative Commons
                Attribution-NonCommercial-NoDerivatives 4.0
                International License (CC BY-NC-ND 4.0)
                https://creativecommons.org/licenses/by-nc-nd/4.0/
```


Raw JSON Data


```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:ng="https://data.ng-london.org.uk/resource/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:crm="http://www.cidoc-
crm.org/sites/default/files/cidoc_crm_v6.2.1-2018April.rdfs#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:cc="https://creativecommons.org/licenses/"
  xml:base="http://data.ng-london.org.uk/resource/" >
  <owl:Ontology/>
  <rdf:Description rdf:nodeID="title_006-001M-0000">
    <rdf:type rdf:resource="crm:E35.Title"/>
    <rdfs:label xml:lang="en">Room 30</rdfs:label>
    <crm:P2.has_type rdf:resource="00A-00DP-0000"/>
  </rdf:Description>
  <rdf:Description rdf:nodeID="subtitle_006-001M-0000">
    <rdf:type rdf:resource="crm:E35.Title"/>
    <rdfs:label xml:lang="en">Spain</rdfs:label>
    <crm:P2.has_type rdf:resource="00A-00DU-0000"/>
  </rdf:Description>
  <rdf:Description rdf:about="000-00A8-0000">
    <crm:P55.has_current_location rdf:resource="006-001M-0000"/>
  </rdf:Description>
  <rdf:Description rdf:about="000-0196-0000">
    <crm:P55.has_current_location rdf:resource="006-001M-0000"/>
  </rdf:Description>
  <rdf:Description rdf:about="000-01AQ-0000">
    <crm:P55.has_current_location rdf:resource="006-001M-0000"/>
  </rdf:Description>
  <rdf:Description rdf:about="000-01AW-0000">
    <crm:P55.has_current_location rdf:resource="006-001M-0000"/>
  </rdf:Description>
  <rdf:Description rdf:about="000-01D6-0000">
    <crm:P55.has_current_location rdf:resource="006-001M-0000"/>
  </rdf:Description>
  <rdf:Description rdf:about="000-02O8-0000">
    <crm:P55.has_current_location rdf:resource="006-001M-0000"/>
  </rdf:Description>
  <rdf:Description rdf:about="000-02OZ-0000">
    <crm:P55.has_current_location rdf:resource="006-001M-0000"/>
  </rdf:Description>
  <rdf:Description rdf:about="000-02PV-0000">
    <crm:P55.has_current_location rdf:resource="006-001M-0000"/>
  </rdf:Description>
```


CIDOC CRM - RDF XML

UK Research and Innovation
Arts and Humanities Research Council

Royal Botanic Garden Edinburgh
The National Gallery
University of Glasgow
V&A
Science Museum Group
Natural History Museum

UK Research and Innovation

Arts and Humanities Research Council

British Library

Royal Botanic Garden Edinburgh

The National Gallery

University of Glasgow

V&A

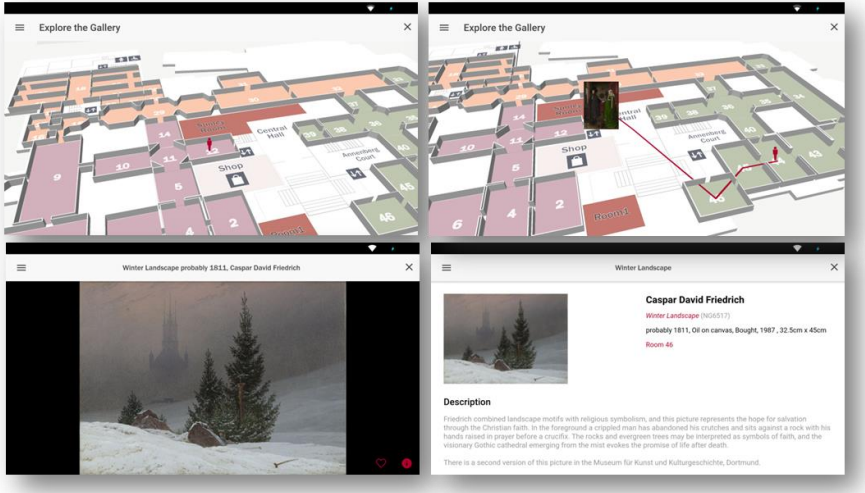
Science Museum Group

Natural History Museum

Re-using National Gallery data.

CrossCult – H2020 EU Project – https://www.crosscult.eu

Data Re-use in practice: Empowering reuse of digital cultural heritage in context-aware crosscuts of European history



Integrated Platform with Dedicated Mobile Apps – Built using re-usable modules

An example research project that dynamically re-used National Gallery data directly via a research API.

The re-use of structured data can lead to multiple benefits

Cross Cult

- Currently store location by room
- Required an efficient tool to record the actual position of a painting on a specific wall.

- Enhanced location data useful for:
- Location recommendation
  - Security
  - Visitor flow
  - Preventive Conservation
  - Documenting exhibitions ….

UK Research and Innovation
Arts and Humanities Research Council

British Library
Royal Botanic Garden Edinburgh
The National Gallery
University of Glasgow
V&A
Science Museum Group
Natural History Museum

UK Research and Innovation

Arts and Humanities Research Council

British Library

Royal Botanic Garden Edinburgh

The National Gallery

University of Glasgow

V&A

Science Museum Group

Natural History Museum

What's next?

What's next? – Linked Open Data

- Explore how best to increase the awareness of NG published PIDs – **F**indable
- Implement a sustainable supportable middleware system, incorporating IIIF and Persistent Identifiers (PIDs), with standard data end-points/APIs – **A**ccessible
- Integrate CIDOC-CRM and external open thesauri and vocabularies (eg Getty Vocabularies, WikiData) mappings and other external Linked Open Data resources – **I**nteroperable
- Assign and publish standard licences for all of our published data – **R**e-usable

UKRI Arts and Humanities Research Council

SSHOC social sciences & humanities open cloud

IPERION HS

E-RIHS

National Gallery: PID System.

Production Local PIDs:

- Create new, independent IDs for all “things” one wants to discuss or describe.
- namespace/AAAA-AAAA-AAAA-AAAA
  - https://data.ng-london.org.uk/0F6J-0001-0000-0000 -> Details of NG35

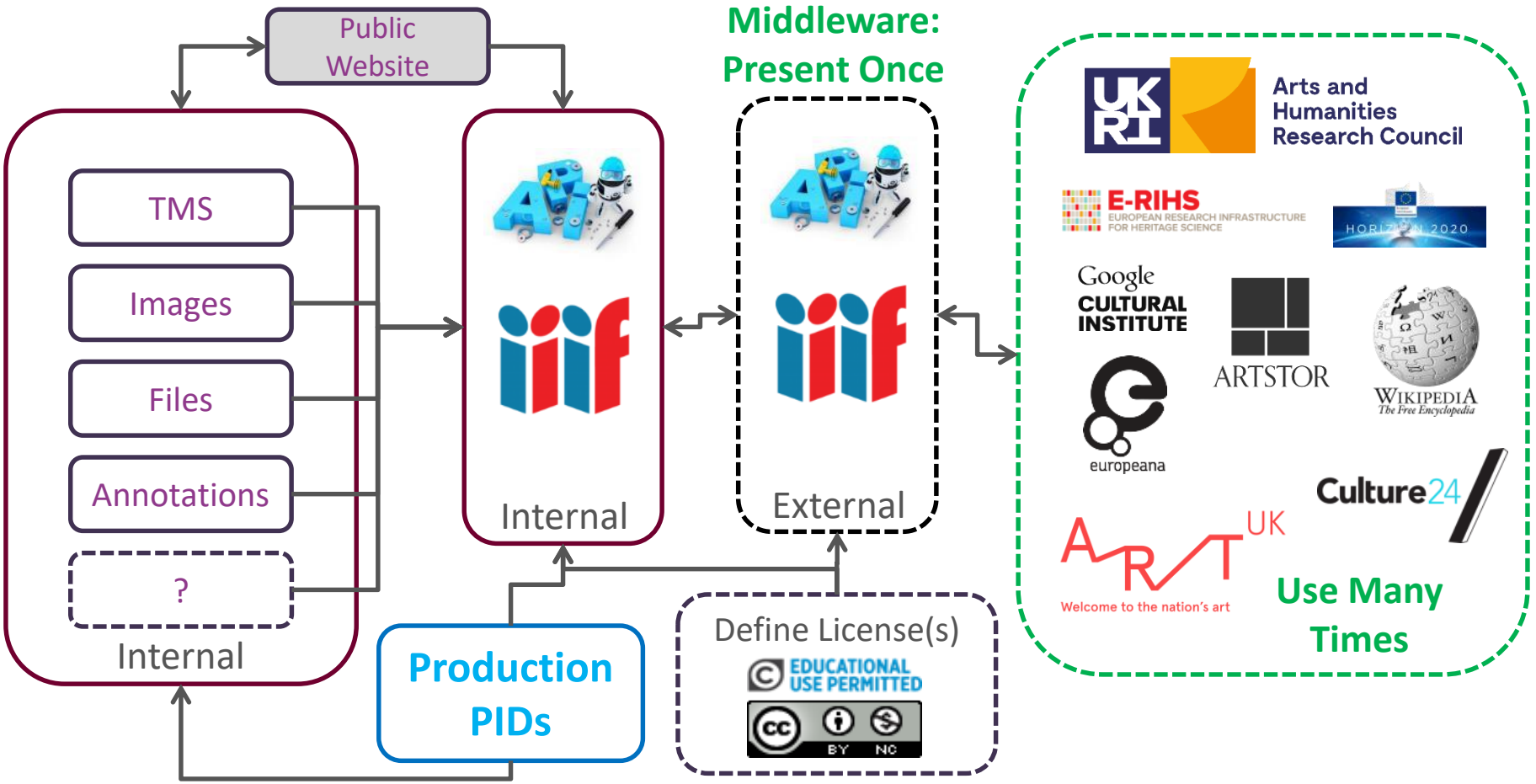
Consider the process of globally registered PIDS:

- Digital Object Identifier (DOI) **doi:10.1186/2041-1480-3-9**
- =handle + added features and rules for persistence, consistency, relationships, semantic interoperability….
- ISO standard, provided by registration agencies, such as **https://datacite.org/**

What will the benefits be from a future combination?

- **Doi:10.XXXX/AAAA-AAAA-AAAA-AAAA**

Building the foundations for a National Collection



UK Research and Innovation

Arts and Humanities Research Council

British Library | Royal Botanic Garden Edinburgh | The National Gallery | University of Glasgow | V&A | Science Museum Group | Natural History Museum

Towards a National Collection: Persistent Identifiers as IRO Infrastructure

Project Launch Webinar, Edinburgh: 6th April 2020

Joseph Padfield (NG)

https://orcid.org/0000-0002-2572-6428

@JoePadfield

Thank you