Towards a vision of an Internet of Cultural Things

Mícheál Mac an Airchinnigh

School of Computer Science and Statistics, University of Dublin, Trinity College, Dublin 2, Ireland

mmaa@cs.tcd.ie

Abstract

The Internet of Cultural Things (IoCT) is a phrase we use to denote all those cultural heritage artefacts which have been assigned their own unique IPv6 address and which are equipped with appropriate wireless communication devices. Each such cultural thing, such as a painting, will be able to express its own ontological identity and communicate with its own "biometric" passport in order to facilitate its exhibition in home museum or abroad. For our own formal backend computing purposes we use a conceptual framework that is an amalgam of top-level ISO standard CIDOC-CRM, realigned with the Sowa-12 framework. Formal encoding is deliberately restricted to OWL2-DL and facilitated by the Protégé 4.0 editor from Stanford. The cultural heritage target is the fine arts of/in Bulgaria, a field with which we have been engaged for several years. The IoCT will facilitate folksonomical tagging of the digitized cultural heritage especially by interested globalized emigrants as well as residents.

Keywords: folksonomy, internet of things, IPv6, ontology, realignment, smart RFID

1. The Conceptual Framework

We choose CIDOC-CRM [1, 2] to be our top level ontology, for the pragmatic reason that it is standard [3]. We use realignment to wed it to the Sowa 12 framework [4]. For folksonomical grounding, recourse is had to Roget's thesaurus [5-7], dating from 1852. The underlying philosophical framework is grounded on the work of Bishop John Wilkins [8], an excellent account of which is provided by Umberto Eco's book on "The Search for the Perfect Language" [9]. Eco concludes (p. 259) that perhaps "the defect in Wilkins' system [was] its prophetic virtue? What if we treated Wilkins as if he were obscurely groping towards a notion for which we have only recently invented a name — *hypertext*?" If indeed Eco's intuition is right, then Roget continues that trend of groping towards hypertext. Many others subsequently contributed to and developed the concept of hypertext that we take for granted today. Of all those, it would be Tim Berners-Lee who would get the ultimate credit through the launching of the World-wide Web by leveraging hypertext [10].

In the context of cultural things we will focus on paintings (drawings and sketches) and photographs. We are wont to attach to these the term ***image*** as used in "digital image, "digitized image", "graven image" [11] and so on. A search in Roget's thesaurus [12] for the concept of image will bring up the entry:

> #17. Similarity -- N. similarity, resemblance, likeness, similitude, semblance; affinity, approximation, parallelism; agreement &c. […] (metaphor) 521; *image* &c. (representation) 554; photograph; close resemblance, striking resemblance, speaking resemblance, faithful likeness, faithful resemblance.

Note how further occurrences (shown underlined) of the term **image** are referred to. So, **image** belongs to (at least) three major categories that are of direct interest: Similarity, Metaphor, and Representation. Note also that since the Roget Thesaurus is online in text form, it is directly amenable to processing and transformation by computer. A simple question comes to mind? How might one fit the thesaurus into a) the CIDOC-CRM top-level ontology, b) an OWL2-DL form? Has it already been attempted? How can we find out?

With respect to the CIDOC-CRM a search of the Erlangen Protégé OWL-DL encoding quickly brings up the class **E38.Image** [which is an E36.Visual_Item] and with which is associated the comment:

> "This class comprises distributions of form, tone and colour that may be found on surfaces such as photos, paintings, prints and sculptures or directly on electronic media. The degree to which variations in the distribution of form and colour affect the identity of an instance of 38 Image depends on a given purpose. The original painting of the Mona Lisa in the Louvre may be said to bear the same instance of E38 Image as reproductions in the form of transparencies, postcards, posters or T-Shirt, even though they may differ in size and carrier and may vary in tone and colour. The images in a "spot the difference" competition are not the same with respect to their context, however similar they may be at first appear. Examples: The front side of all 20 Frs notes, the image depicted on all reproductions of the Mona Lisa." [13]

Now let us imagine (with difficulty?) that the name "Mona Lisa" is not known to the "reader"? For emphasis let us substitute another name: "Kyustendil Woman" (Figure 1) in her place? How does this effect the above explanatory text? Clearly, the location of the physical painting is very different. It is known, but not by many, to be found (usually) in the Regional Art Gallery in Smolyan, Bulgaria.
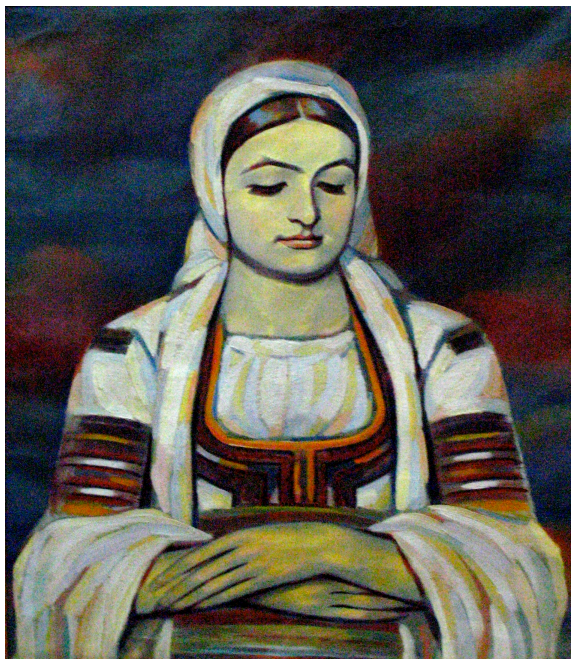

Figure 1. Жена от Кюстендилско

The title of the painting translates as "Woman from Kyustendil region". We now have **two** English titles and **one** Bulgarian title. However, postcards of the image are still being sold in Smolyan. On the reverse of same, the (older?) Bulgarian title is "Селянка от Кюстендилско" (Peasant woman of Kyustendil region). That now makes **three** titles in English and **two** in Bulgarian for the one painting.

From a different perspective, one who uses, for example, WordNet, will discover the concept of image in another way. Of all those senses currently listed therein, the one closest to what we have in mind here is

> noun: A representation of a person (especially in the form of sculpture); "the coin bears an effigy of Lincoln"; "the emperor's tomb had his image carved in stone" [14].

Pragmatically, in our research work everything is being encoded in OWL2 [15] the Semantic Web Ontology Language, restricted to the description logic level (DL). For this purpose we use Protégé-OWL version 4.0, the "Ontology Editor for the Semantic Web" [16]. In practice: We use the top-level CIDOC-CRM ontology from Erlangen [13] accessed through the Protégé Editor version 4.0 [17]. However, such stuff is backend computing, behind the scenes, so to speak. The front of stage is for the benefit and pleasure of the audience, the masses, those who are simply curious or interested or even passionate about "their" art. This is the realm of the educated, already covered by the (formal) language of the field, and more importantly in the Internet of Cultural Things, by the majority of those who belong to the culture of which the art is representative. The language of the latter is covered by the term folksonomy [18, 19]. One way in which "all the connected" can participate might be to use Hyperwords [20] with the Firefox browser. Unfortunately, there is no translation to Bulgarian!

Figure 2. Sofia University

2. Populating

From a pragmatic perspective it is the *people* who will determine whether or not the digital culture will become an authentic extension of the existing cultural heritage. How will the people determine the development? Here we give three possible insights as "examples".

Example 1: Galaxy Zoo [21, 22]: (around 200,000) users tag images of galaxies; here the mass intelligence of the human swarm comes into play; no special astronomical knowledge is needed. One imagines how the images of galaxies might be replaced by images of paintings. Do we not need a *Fine Art Zoo*?

Example 2: Google Earth [23]: users tag the earth with names (in their own languages) and images (usually digital photographs). The latter are often misplaced, in the sense of not being precise with respect to location. Moreover, there is usually no indication as to the direction and orientation of the photographer who points the camera to shoot the scene.

In Google Earth one gets to a place by "flying", recalling (perhaps inadvertently) the manner of travel in Second Life. For example, to get to Sofia, Bulgaria is easy. To get to the University of Sofia requires a little more effort. That is to say, by "flying", many possibilities are offered. One needs to know something of the local geography and the academic world. In this case, it turns out that the University of Reading (UK) has already put down an accurate place marker entitled "Софийски университет" (Sofia University). I know that this is the right place. I also know that this is where the Faculty of Philosophy is located and that the Faculty of Mathematics lies elsewhere. To personalize my view I have tagged it.

From Google Earth I can go directly to Google Maps and print out the result, or in the spirit of the digital age, grab a screenshot (Figure 2). That is to say I can show you exactly what I have been talking about via a specific image. I could also have included a video clip of how I achieved this task. Finally, it is worth noting that it is possible to associate a Wikipedia article with a particular place. Instructions on how to do this are available at Frank Taylor's Google Earth Blog [24].

Example 3: Thought Lab at Europeana [25]: this is a research prototype of a semantic search engine. It is of considerable importance for our research precisely because it attempts to negotiate the retrieval of a search with the aid of 4 very different types of thesauri: Joconda [26] of French origin, IconClass [27] of Dutch origin, RKD Artists [28] also of Dutch origin, and the AAT [29] from Getty. A simple semantic search using "woman" brings up some interesting results; a search for "Kyustendil" also gives a good result with geographical description; search for "Kyustendil" + "Woman" is futile, as expected. Further exploration behind the scene reveals a Europeana datacloud [30] of which WordNet FR (Wolf), WordNet UK (Princeton) and WordNet NL (Cornetto) are also key components. Based on an exploration of the semantic search engine one can see how it might be extended to embrace Bulgarian Art. One surmises clearly what has to be done for the Bulgarian Semantic Web and, in our general context, the integration of the Bulgarian WordNet [31]. A good starting point might be the Catalogue of the Paintings in the City Art Gallery of Plovdiv [32] since the information is concise and presented in parallel Bulgarian and English texts.

3. Internet of Things

The Internet of Things (IoT) [33] is an evocative phrase deliberately designed to conjure up an interconnected wireless world of things, each with their own unique IPv6 address [34] that embraces animate objects such as "cows in the field" [35] and inanimate objects such as books in a library, or paintings in a museum. For the wireless communication, two complementing technologies are in the frame: smart RFID tags [36] and Near Field Communication (NFC) devices [37]. In a recent public response document to the EU Commission [38], we proposed that one ought to consider how the new Internet technologies might be deployed in the general area of cultural heritage, with a specific focus on the fine arts. Philosophically we presented a case that would move the debate forward from the usual "application suspects" [39] of commerce, health, energy, environment and so on, to the domain of Cultural Heritage. To argue the case we presented a few scenarios, briefly summarized in the following paragraph.

We identified 4 distinct types of scenario that might be used in demonstrator projects for the purpose of rolling out the Internet of Things in a way that might be picked up and appreciated by the wider public of the EU. The first scenario (with 2 subdivisions) concerned the use of the IoT technologies for facilitating the usual transactions involved in the organization of (conventional) art exhibitions. The second scenario (with 2 subdivisions) extended the conventional into the digital world of art. Scenario 3 (with 2 subdivisions) advocated IoT enabling technologies for the conventional Art Gallery. It is here that use of the NFC technology can be exhibited. The final scenario is, in a sense, the odd one out. Instead of paintings in an Art Gallery we have books in a library. The speculated use of the IoT technology to identify theft or damage of rare (one copy) books applies, inter alia, to art works in general. Since the publication of our detailed response on the Internet of Cultural Things, we have been working on another practical case.

Let us imagine that we are called upon to develop a practical scenario that incorporates the IoT smart RFID technology into the existing Art Gallery in Smolyan wherein hangs the painting of "Kyustendil Woman" (Figure 2)? For brevity we will use the tag "wink" (мигане) to denote a smart RFID chip. In the first immediate stage we propose the attachment of 4 winks: W1) on the reverse side of the canvas containing the unique IPv6 address of the art work; W2) on the reverse side of the frame; W3) on the wall label identifying the art work; W4) on the wall above the art work identifying the physical space/location.

W1 is recorded on both the digital passport and the digital dossier of the art work. If the separation distance between W1 and W2 exceeds a given threshold an alarm event is generated signalling that the canvas and the frame have become physically separated. The event is valid if the painting needs to be taken out of the frame for cleaning, restoring, X-ray examination and so on. Otherwise the alarm indicates potential theft! If the separation between W2 and W4 exceeds the threshold then the alarm event indicates that the painting and frame have been removed from the wall. This would positively be taken to mean preparation for an exhibition elsewhere in the gallery or to another "foreign" gallery. The wall label W3 is intended for the new digital installation which will

provide the descriptive text in the language of the viewer through a personal mobile device.

Figure 3. "Kyustendil Woman" in Regional Art Gallery, Smolyan

In Bulgaria every(?) painting has its own dossier and passport. The passport usually contains a photograph of the painting, the name of the gallery to which it belongs and other essential identifying information such as one might find in a catalog entry. The dossier contains the complete record of the painting, its provenance, the fund used in its acquisition, and so on. One can easily imagine that every "significant" painting in the EU will "eventually" have its own biometric passport. Such a passport might be prototyped similar to the current vCard that is in widespread use [40]. A facsimile of the lead page of a typical paper passport is shown in Fig. 4 (courtesy of the Plovdiv Art Gallery). From the corresponding catalogue of the gallery already referred to earlier we know that it is "oil on canvas", size 97x53 cm, is "signed on bottom left" and has inventory number 653. In the digitization of the painting, the passport, and the dossier, one will wish to retain backwards compatibility. Specifically, one will note that the inventory number has been changed from 1453 at some earlier time to its current 653. Such changes always occur. The painting might subsequently be "sold" to another gallery. But the provenance data must be retained bother by seller and then by buyer. This is probably the way it has been in the past. The IoCT future anticipates the tradition will continue electronically and automatically.

Figure 4. "Before a Storm", Grigor Savov 1942, in City Art Gallery, Plovdiv

A painting provided with its set of smart (i.e.) active RFID chips as described above for the purpose of being exhibited will also be open to the possibilities of being monitored in any environment in which it finds itself. In other words, once it becomes a member of the Internet of Cultural Things then its "health" may be checked regulary, whether it is in storage, in transit, or on exhibition.

4. Things to be done

There are terminological (vocabulary) frameworks already in use. There are no good reasons for all to adhere to any single one. Legacy systems must always be accommodated or realigned with any *de facto* or official standard. It is a well-founded hypothesis (of computing) that there will always be progress in the computer codes. We use the term realignment instead of the more common alignment to emphasis the recursive necessity by which we are always bound. A simple search with terms CIDOC + alignment + "Getty vocabularies" will unearth the research activity since 2006. It is interesting to note that one recent paper on the subject has the title "Principles for Knowledge Engineering on the Web" [41].

References

1. W en. *CIDOC Conceptual Reference Model.* [cited 2009 Mar 6]; Available from: http://en.wikipedia.org/wiki/CIDOC_Conceptual_Reference_Model.
2. Nick Crofts, et al. *CIDOC Conceptual Reference Model, version 4.2.2.* 2008 [cited 2008 Jan 27]; Available from: http://cidoc.ics.forth.gr/.
3. ISO 21127:2006. *Information and documentation -- A reference ontology for the interchange of cultural heritage information.* 2006 [cited 2009 Mar 6]; Available from: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34424.
4. Sowa, J.F., *Knowledge representation : logical, philosophical, and computational foundations.* 2000, Pacific Grove, CA ; London: Brooks/Cole. xiv, 594.
5. Roget, P.M. *Roget's Thesaurus (1911).* [cited 2009 Mar 8]; Available from: http://machaut.uchicago.edu/rogets.
6. W en. *Roget's Thesaurus.* [cited 2009 Mar 8]; Available from: http://en.wikipedia.org/wiki/Roget's_Thesaurus.
7. Roget, P.M., *Roget's Thesaurus,* ed. S.M. Lloyd. [1852] 1962, 1982, Burnt Mill, Harlow, Essex: Longman Group Limited.
8. W en. *John Wilkins.* [cited 2009 Mar 9]; Available from: http://en.wikipedia.org/wiki/John_Wilkins.
9. Eco, U., *The search for the perfect language.* The making of Europe. 1995, Oxford, UK ; Cambridge, Mass., USA: Blackwell. x, 385 p.
10. W en. *Hypertext.* [cited 2009 Apr 23]; Available from: http://en.wikipedia.org/wiki/Hypertext.
11. Mac an Airchinnigh, M. (2005) *The Graven Image — Digitized and Philosophized.* . Review of the National Center for Digitization **6**, 12–44.
12. Roget, P.M. *Roget's Thesaurus (1911).* 2004 [cited 2009 Mar 16]; eBook #10681:[Available from: http://www.gutenberg.org/files/10681/10681-h-body-pos.htm.
13. Schiemann, B., M. Oischinger, and G. Görz *Erlangen CRM Version 2009-03-30.* 2009 [cited 2009 Apr 15]; Available from: http://www8.informatik.uni-erlangen.de/IMMD8/Services/cidoc-crm/erlangen-crm_090330_5_0_1.owl.
14. WordNet. *WordNet.* 2008 [cited 2008 Jan 27]; Available from: http://wordnet.princeton.edu/.
15. W3C. *OWL2 Web Ontology Language — Direct Semantics, W3C Working Draft 21 April 2009.* 2009 [cited 2009 Apr 22]; Available from: http://www.w3.org/TR/owl2-semantics/.
16. Protégé OWL. *Ontology Editor for the Semantic Web.* [cited 2009 March 6]; Available from: http://protege.stanford.edu/plugins/owl/download.html.
17. *Protégé OWL Editor.* [cited 2009 Mar 25]; Available from: http://protege.stanford.edu/overview/protege-owl.html.
18. Mac an Airchinnigh, M. and K. Sotirova, *Keyimage Ontologization & Folksonomy in Web 2.0 Social Space,* in *Değişen Dünyada Bilgi Yönetimi Sempozyumu,* S. Kurbanoğlu, Y. Tonta, and U. Al, Editors. 2007, H.Ü. Bilgi ve Belge Yönetimi Bölümü: Ankara. p. 32-41.

19.   W en. *Folksonomy*.   [cited 2007 Jan 4]; Available from:
        http://en.wikipedia.org/wiki/Folksonomy.
20.   W en. *Hyperwords*.   [cited 2009 Apr 23]; Available from:
        http://en.wikipedia.org/wiki/Hyperwords.
21.   Charman-Anderson, S. *A galaxy of helpful people*. 2009 [cited 2009 Mar
        9]; Available from:
        http://www.guardian.co.uk/technology/2009/jan/15/internet-
        astronomy.
22.   Lintott, C. *Galaxy Zoo*.   [cited 2009 Mar 9]; Available from:
        https://galaxyzoo.org/.
23.   Technophile. *Google Earth*. 2009 [cited 2009 Mar 9]; Available from:
        http://www.guardian.co.uk/technology/2009/feb/12/google-earth-
        technophile.
24.   Taylor, F. *Google Earth Blog*.   [cited 2009 Mar 9]; Available from:
        http://www.gearthblog.com/blog/archives/2006/12/multilingual_wikip
        ed.html.
25.   Europeana. *Thought Lab, a research prototype of Europeana's semantic
        search engine*.   [cited 2009 Apr 24]; Available from:
        http://www.europeana.eu/portal/thought-lab.html.
26.   Europeana. *Thesaurus Joconda*.   [cited 2009 Apr 23]; Available from:
        http://www.culture.gouv.fr/documentation/joconde/fr/recherche/aide-
        icono.htm.
27.   Europeana. *Thesaurus IconClass*.   [cited 2009 Apr 23]; Available from:
        http://www.iconclass.nl/libertas/ic?style=index.xsl.
28.   Europeana. *Thesaurus RKD Artists*.   [cited 2009 Apr 23]; Available from:
        http://english.rkd.nl/.
29.   Europeana. *Thesaurus Art and Architecture (Getty)*.   2009]; Apr 23].
        Available from:
        http://www.getty.edu/research/conducting_research/vocabularies/aat/.
30.   Europeana. *Europeana Datacloud*.   [cited 2009 Apr 23]; Available from:
        http://eculture.cs.vu.nl/europeana/www/datacloud.html.
31.   *WordNet BG*.   [cited 2009 Apr 23]; Available from:
        http://catalog.elra.info/product_info.php?products_id=802.
32.   Градска художествена галерия Пловдив, *Каталог живопис*. 2002,
        Пловдив: ЛЕТЕРА.
33.   W en. *Internet of Things*.   [cited 2008 Nov 14]; Available from:
        http://en.wikipedia.org/wiki/Internet_of_Things.
34.   W en. *IPv6*.   [cited 2008 Nov 14]; Available from:
        http://en.wikipedia.org/wiki/IPv6.
35.   Dodson, S. *The net shapes up to get physical*. 2008 [cited 2008 Nov 15];
        Available from:
        http://www.guardian.co.uk/technology/2008/oct/16/internet-of-things-
        ipv6.
36.   W en. *Radio-frequency identification*.   [cited 2008 Nov 14]; Available
        from: http://en.wikipedia.org/wiki/Radio-frequency_identification.
37.   W en. *Near Field Communication*.   [cited 2008 Nov 14]; Available from:
        http://en.wikipedia.org/wiki/Near_Field_Communication.
38.   Mac an Airchinnigh, M. *Towards a vision of an Internet of Cultural Things*.
        [Position Paper] 2009; Available from:

http://ec.europa.eu/information_society/policy/rfid/library/index_en.ht m#iotcons.

39. Commission Staff. *Early Challenges regarding the "Internet of Things".* [cited 2008 Nov 14]; Available from: http://ec.europa.eu/yourvoice/ipm/forms/dispatch?form=IOTconsultati on.

40. W en. *vCard.* [cited 2009 Mar 25]; Available from: http://en.wikipedia.org/wiki/VCard.

41. Schreiber, G. and L. Aroyo, *Principles for Knowledge Engineering on the Web*, in *Association for the Advancement of Artificial Intelligence (www.aaai.org)*. 2008.