

IDENTIFICATION TOOL FOR CANCELLATIONS OF THE OTTOMAN EMPIRE

George I. Stassinopoulos
School of Electrical Engineering and Computer Science
National Technical University of Athens
Zographou Campus, 157 73
Athens
Greece
stassin@cs.ntua.gr

Abstract

The OCIT (Ottoman Cancellations Identification Tool) places partially preserved cancellations on Ottoman stamps within the prestigious “Cancellations of the Ottoman Empire” as reported by prominent scholars. It also serves as a complete electronic index of major publications in this area, each having different formats and conventions for identifying and listing. Over 6500 Ottoman cancellations from more than 1800 sites of the Ottoman Empire in the Balkans, Near & Middle East are included. Although a complete development by itself, OCIT is taken as a first step for future extensions for integrating collections of different items under common criteria and a variety of scientific objectives. Key problems encountered are reported and functional extensions and generalization of scope are suggested. This aims at a generic indexing and cataloguing tool for cultural heritage collections. Fragments have to be mutually identified as being instances of the same prototype (the die used), which however is unknown. It is manifest able only through partial, hopefully partly overlapping strikes. Query constructs in common use, like wildcards, are not sufficient. Special emphasis is given to metadata annotations and links to historical events and geographic / chronological assignment, consistency in *distributed use* and *retrospective* structural updates under collaborative control.

INTRODUCTION

We present in a bottom up fashion experience and upcoming challenges for ‘digital preservation of cultural heritage’ activities involving interested groups in the wider public. As information technology awareness and skills penetrate collectors, enthusiasts and hobbyists, a new potential for wide scale collaborative projects in

documentation, research and ultimately preservation opens up. Such projects can be driven by a wide range of motivations, from pure scientific interest and satisfaction in research, through intense collectors' drive, to material profit. Without geographical barriers and real time constraints the potential appears extremely promising with large number of skilled individuals drawn into quite extended and far reaching activities. Hence we lie, so to speak, in the cross-section of 'the long tail' ([8]), digitalization of cultural items ([9]), and 'collecting / hobby / amateur' research activities. Moreover the application discussed can serve as a model in a wider sense. It addresses not cultural items per se, but rather their manifestations of differing integrity and quality widely distributed across the public. Cancellations on stamps and envelopes, seals, coins and similar items circulating in thousands as 'strikes' or 'prints' of lost or extremely rare one-off 'dies' are affordable and widely distributed. These items constitute nonetheless important holders of cultural, historical and artistic content and mobilization of their collectors should be promoted via information technology tools. The 'content' consists then of digital records held in individual 'data bases'.

The paper discusses key issues necessary to be resolved in the area of distributed and collaborative documentation in distributed and collectively designed 'data bases'. User friendliness and a realistic direct approach commensurate to each area's scope is essential, if one targets a wide dissemination and extended use not only of the actual results, but also of the use and evolution of corresponding documentation tools. We first present in some detail the existing application. We then draw lessons and formulate guidelines for an exposure in a distributed pier to pier environment. Some key technical decisions are finally presented. These are intended to support the view that such a path is indeed possible in today's information and networking environment.

WORKED OUT APPLICATION

We describe the flavor, scope, extent and use of the developed application after a brief introduction of the application domain and the interest therein.

The Domain

Ottoman Cancellations are particularly interesting mainly among philatelists. These cover a wide geographical area and involve multinational, multilingual and multicultural toponyms (post office geographical names), frequently changing over the declining period (~1863 - 1922). This was also the formative period for a number of nation states in Southeastern Europe, Middle East and North Africa, which brings national and regional aspects into the fore light. Moreover their poor official documentation, resulting into ever appearing new findings and surprises, as well as rarities and forgeries add to the excitement offered by this kind of collection.

Bilingualism or rather the use of two alphabets, Ottoman/Turkish (Arabic) and French (Latin) is of prime importance. Cancellations were originally in Arabic. This is traditionally known as the 'Brandt period' and separately documented as will be explained below. Subsequently bilingual cancels appeared mainly in circular form. Arabic appeared mostly at the top, French at the bottom. Different Arabic calligraphic types mainly rika and later nasx, were used. Rika is a form of Arabic stenography allowing fast but at the same time aesthetically appealing handwriting. It is of Turkish origin and was widely used in the Ottoman administration [6], hence also on the cancels produced by the Ottoman Post and distributed throughout the Empire.

The application described in this work involves fragments of the strikes of cancellations on stamps, envelope fragments or entire envelopes. The cancellation appears more often than not only partially. Entire cancellations with clearly struck fields are relatively rare and sometimes extremely expensive. Hence we are particularly interested in difficult to read, partially preserved and unclear cancellations. These are numerous and relatively affordable and the collector's scope and satisfaction are increased, if he is able to acquire, recognise and handle a large amount of such samples. If the bottom part is missing, only the Arabic script is available, sometimes also partially and / or badly readable. If the left (right) part is missing, the start (end) of the French together with the end (start) of the Arabic rendering of the post office name is readable. Often these two names are different, e.g. 'DAMAS' (Damascus) in French was officially termed as شام (Sham) in Ottoman times. Hence a surviving right edge of a cancellation would allude to a place written in Arabic with a starting 'ش' (shin) and in 'French' with an ending Latin 'S' – a

difficult riddle to the uninitiated. The text based search incorporated into OCIT (see Fig.7 below) is fully capable to pin down the particular candidate cancellations in this and similar cases. Over 6500 distinct cancellations from more than 1800 sites of the Ottoman Empire are included in the application described below. Within this quantitative range, the lists of both cancellations and sites are open ended. The corresponding literature consists of four major reference listings ([1] – [4]) as well as maps [5]. References [2] and [3] draw also from official post office records, which are however incomplete due to the gradual disintegration and loose control exercised over a vast geographical area. Facing the Arabic text, one encounters post office names (frequently differing from toponyms on modern maps) and / or common expressions of geographic, administrative or cultural scope. Post office names have to be represented at least four fold, see Fig.1. The first ('Ottoman', i.e. in Arabic) and second ('Latinised', i.e. in 'French') columns are the one likely to be found on the cancellation itself, more often than not altered in spelling. The second is not necessarily the modern name, even for locations in modern Turkey. Thus column 3 is essential in rendering the post office location name, as used today, in each and every country and taking into account numerous changes due to historical, cultural and national sensitivities and a variety of other reasons. This would be the name found by an air traveller buying from the airport an internationally edited map of its destination country in the region in question. A locally edited map of the same country would print the same name in the native language and alphabet also provided by OCIT. There is however more. A loose list of further names has to be included encompassing all those names, in whatever language, as used by former 'Ottoman citizens' of various nationalities for the place in question. Take 'Athens' as an extreme case. This is not actual Athens, which was no part of the Ottoman Empire at the period of concern, but rather a relatively small locality in the North Eastern Black Sea coast of Turkey (Pontos). It was founded and colonised by Perikles himself on the 5th century B.C. and appropriately named after Athens. This name prevailed up to Ottoman times rendered in Arabic as *اتنه* (in modern Turkish script *Atına*). It is still recognizable by the older generation. Nowadays it is officially known as Pazar, a name never used in Ottoman cancellations of this locality. However the Russian misspelling *АТИА* is indeed found on one (Ottoman!) cancellation, probably a remnant of the brief occupation by the Russian army in 1915. We have touched during this description

upon the interrelationship of geographic names with ‘*cultural and political groups*’ as understood in the Getty TGN ® ([11], par 1.1.3.4.2). Although a difficult issue, the importance of presenting in parallel different names in different languages bearing different emotions to different cultures and people, all for the same place cannot be overlooked. It is a case where multilingualism and dates meet at the same toponym featuring on different cancellations.

What has been presented so far is not a way of cataloguing the cancellations themselves, but only toponyms of post offices issuing a particular set of cancellations. This set consists of differently spelled renderings of either or both of columns 1 and 2, according to the particular dates. These combinations are mapped onto different shapes and sizes. As said, post office name entries as in Fig.1 are of the order of 1800, while cancellations are more than 6500 in total. Localities can be small with just one post office. Major centers, e.g. the capital, appear in different names: Der Saadet (‘Gate of Happiness’), Der Aliye (‘Sublime Porte’), Constantinople, Stamboul, İstanbul, all successive Ottoman designations using Farsi (Der), Arabic (Saadet, Aliye), Latin / Greek (Constantine / polis) and versions in Turkish (İstanbul, Stamboul). Additionally the City itself has to be split into individual districts (Galata, Pera, Arsenal, Tophane, etc.), each with its proper post office issuing over time tens of different cancellations.

Ottoman	Turkish	Latinised	Multilingual
آينه روز	Ayanoroz	Aghion Oros	Ἄγιον Όρος
ارضروم	Erzurum	Erzurum	Θεοδοσιούπολις
غولوس	Golos	Volos	Βόλος
يانيه	Yanya	Ioannina	Ἰωάννινα
كره بنه	Gerebine	Grevena	Γρεβενά
قدس شريف	Kuds-i Şerif	Jerusalem	ירושלם
بيت لحم	Beyt ül-Lahm	Bethlehem	بيت لحم
اتنه	Atına	Pazar	Αθήναι, АТИА
مناستر	Manastir	Bitola	Битола, Μοναστήριον
اسكوب	Üsküb	Skopje	Скопље
دراج	Drac	Durrës	Δυρράχιον
قلبه	Filibe	Plovdiv	Πλοβдив, Φιλιππούπολις
قره حصار صاحب	Karahisar Sahib	Afyon Karahisar	Ἄκροϊνόν

Fig.1. Post Office Multilingual / Multicultural Rendering

The situation is somewhat simpler with common expressions as in Fig. 2 below. These are sometimes accompanying the post office name renderings as above in various positions and combinations. Only the first columns actually appears (plus easy to handle equivalents in French) and columns 2 & 3 serve only for revealing the spelling and the meaning of each line. There are about 100 such entries.

Ottoman	Turkish	Meaning
خانه	hane	office
پوسته	posta	post
شعبه	şube	section
اسکله	iskele	embankment
قرق	kırk	fourty
یول	yol	road
چشمه	çeşme	fountain
اداره	idare	direction
واپور	vapor	steamer

Fig.2. Some Common Expression on Ottoman Cancellations

The Ottoman Cancellations Identification Tool

The OCIT implemented as a stand alone Windows application is a registration, identification and search tool for Ottoman Cancellations. Each cancellation record contains the exact spelling in Arabic and / or French as applicable, shape and size code, color(s) of strike, page and number as referenced in the literature [1] – [4], presence and placement of common expressions, link to characteristic image files and association to the post office. Post office records contain all possible names as explained in Fig.1 and are associated to former vilajets (large Ottoman administrative regions) and present date countries. The main form is depicted in Fig. 3. Post office names, vilajets and countries can be selected, entered and queried either in Arabic with names used in Ottoman times or in the present language and alphabet, as applicable today (Turkish, Greek, Albanian, Slavic languages, Arabic). As soon as a

geographical entity (the whole Empire, a vilayet, a country, or a specific post office) is selected, all cancellations appear on the main form.

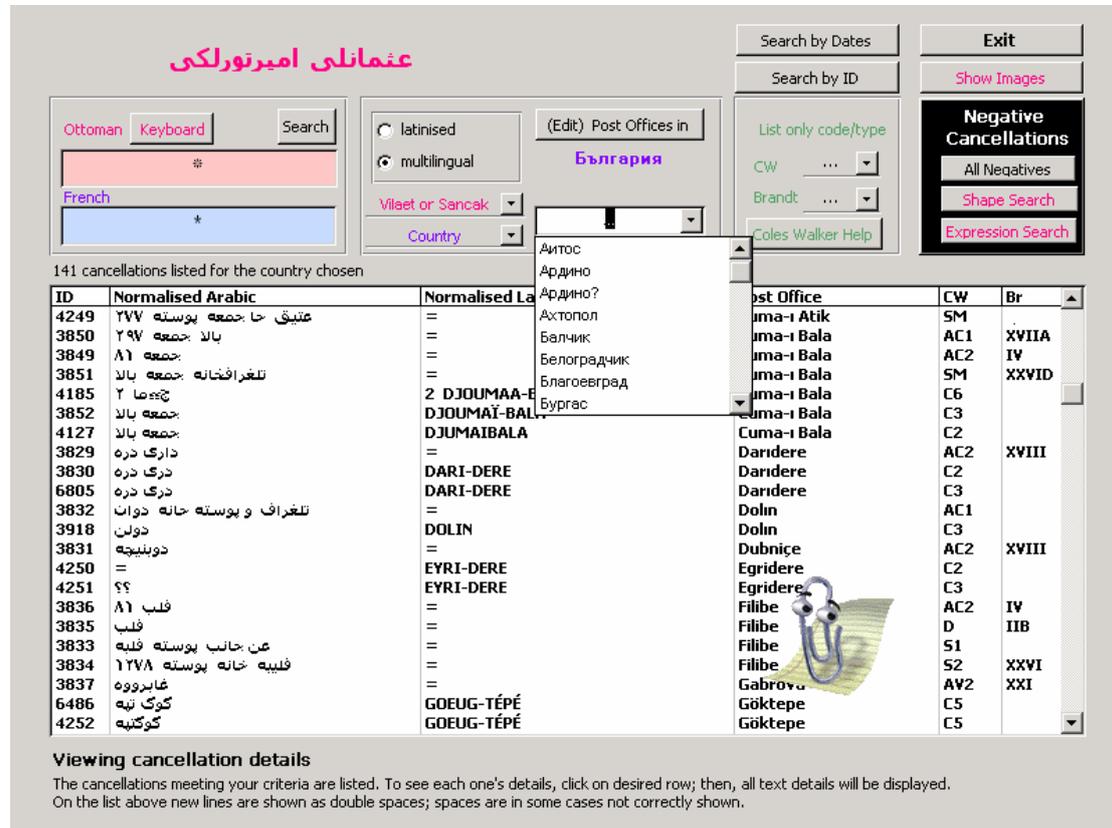


Fig.3. OCIT Main Form

So far OCIT can be seen as a mere indexing of the data found in [1] – [4]. The main value of this tool lies however in the ability to identify fragmented or partially readable items. This search can be based on

(a) *shape* according to various coding conventions used in the literature as well as specific OCIT provided simplified characteristics and size selection, see Fig.4,

(b) *location* of common expressions, color, see Fig.5, which is particularly helpful in the so-called ‘negative cancels’ with expressions and post office name entangled in two dimensions, according to the space available and the calligraphic aspirations of the engraver. A color code distinguishes the various common expressions (always in white on actual negative cancels) and matches pop up as thumbnails, see Fig. 6.

(c) text appearing on the cancellation, i.e. Arabic and/or Latin characters as far as readable. Wildcard characters can be used in the query, matching however proceeds along a number of alternative readings of a normalized text resident in the data base.

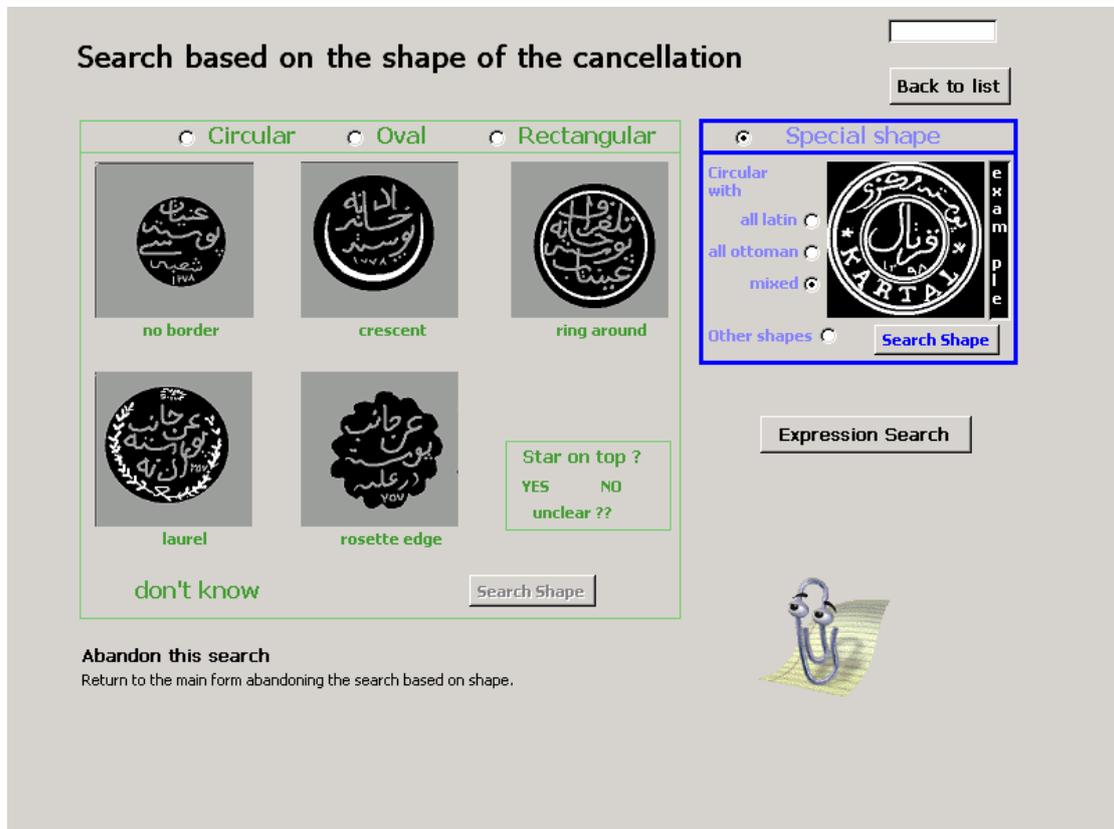


Fig.4. OCIT Shape Based Search

Search based on the placement of Ottoman Post expressions on the cancellation

[Back to list](#)

an canib

top
 middle
 bottom
 unclear

nonexistent

telegraf

top
 middle
 bottom
 unclear

nonexistent

posta

top
 middle
 bottom
 unclear

nonexistent

şubesi

top
 middle
 bottom
 unclear

nonexistent

hane(si)

top
 middle
 bottom
 unclear

nonexistent

Numerals

hindi
 arabic

Search taking also into account the hindi numeral on a negative cancellation
 Optionally enter clearly identifiable numerals using "*" for unclear ones (e.g. 1*04, if second character unclear).
 Put number (above) in hindi or (below) in arabic numbers).

Fig.5. OCIT Common Expression Location Based Search

558 cancellations with image(s) found

[Previous](#) Page 23 of 47 [Next](#) [First](#) [Middle](#) [Last](#) [Back](#)

 Isparta hane 1278 posta 894	 ujretli alınmıştır P.P. 2346	 931	 an canibi posta 253 Izmir 937
 hane Izmir posta 938	 Izmir 304 940	 Izmir posta telegraf ve hane 1313 5534	 Izmir posta telegraf ve hane 1315 939
 Izmir hane si Merkezi posta 328 942	 Pazarı İzmir posta şubesi 941	 Joban postası si şube 3415	 Zopan postası si şube 3425

Fig.6. Thumbnail Presentation of Negative Cancels

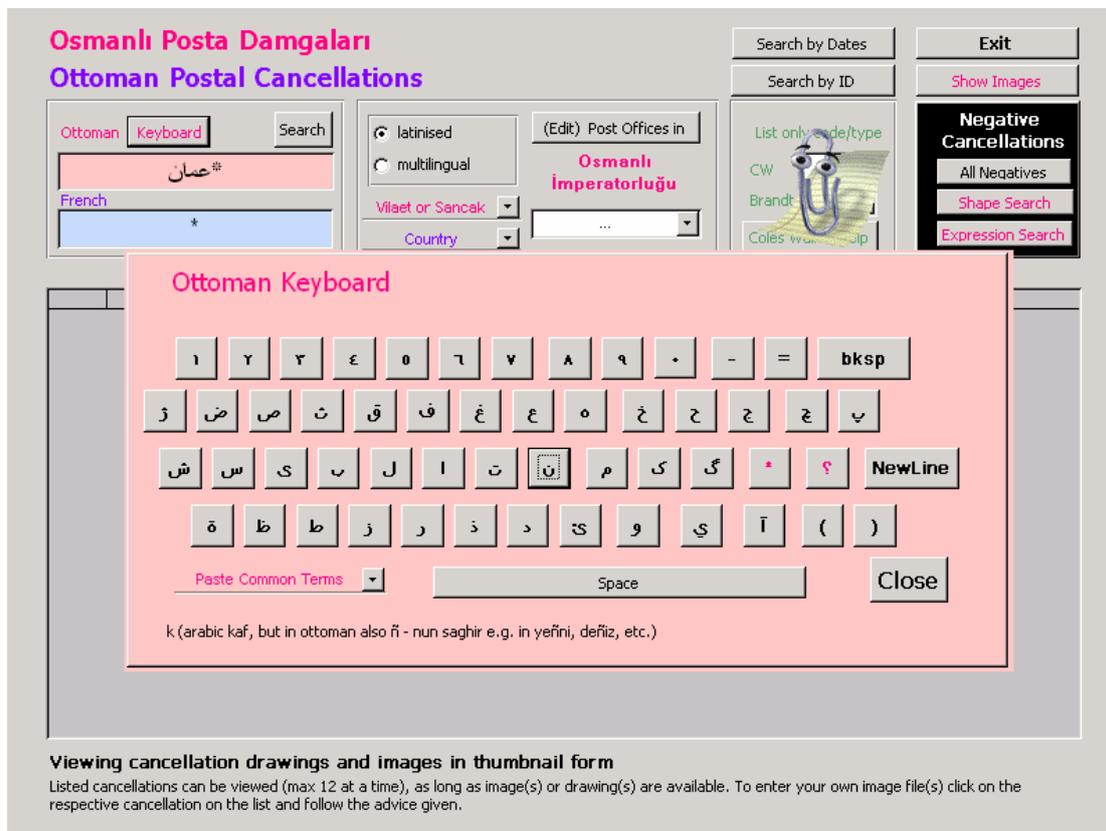


Fig.7. Ottoman Text Based Search (upper left) via embedded ‘Ottoman Keyboard’

Unicode is indeed valid across platforms and different settings. However software keyboards for different languages are still a source for confusion and disappointment even for experienced users. For that purpose an Arabic keyboard had to be embedded into the application, as shown in Fig.7.

A search with *ش (right-to-left) in the Ottoman and *S (left-to-right) in the French field, is now able to pin down cancellations in Damascus. The text search, aided by spelling variations embedded in the Ottoman and Arabic rendering of toponyms constitutes a real add on to the conventional search in printed catalogues. Lexicographic listing, as done in the literature, is extremely vague. Geographic is no much help either, given the large number of relatively unknown localities as well as the proliferation of multiple uses of common names across the vast extents of the empire, e.g. place names like Akşehir (‘white city’).

GENERALIZATION OF SCOPE

OCIT, as described, can be seen as an instance of a class of cultural applications with the following generic characteristics.

Ease of Use / Pragmatic Approach

The degree of information technology penetration and use should be constantly evaluated with consideration of trade-offs against expected gains. In the case of OCIT, a constant issue coming up on each presentation is the involvement of OCR and automated identification via image processing. This is a typical case of technology centric approach, which could easily annihilate the main assets, namely ‘ease of use’, ‘willingness to adopt and use’, cost effective and timely deployment given the prerequisites and aspirations of the target user group. The trade-off is between development time and cost for a feature attacking a particularly difficult and hitherto unexplored domain, i.e. fragmented, misprinted text on a circular or even a chaotic two dimensional set up. Results would, at best, only be reliable for easy to handle cases, i.e. precisely for those cases where OCIT is superfluous. On the other hand purely textual search taking into account ambiguities in place name renderings of a whole region, under different languages and scripts, is an issue central to ‘Ottoman Cancellations’ as well as a methodology useful in a general. This is pursued in the next paragraph.

Equivalent Perception

The scope is a collection of items each characterised as being a strike of a particular ‘die’. This is the canceller in the case of cancellations, the die in the case of coins. Cancellations are no part of ‘museum items’ the cancellers themselves might be, but area largely lost forever. In the case of numismatics both coins and their die(s) (extremely rare) *can be* museum items. In that respect and in view of the large number of strikes of different integrity and preservation quality, a framework like [9] is only partially applicable. As a rule there is no access to the ‘die’ itself and the strikes at hand are imperfect and/or partial images of it. The textual content of the ‘die’ is often

partially known and/or deducible with a degree of uncertainty, due to imperfect strikes, damage, wear etc. In some cases, i.e. in numismatics, the ‘die’ itself is not unique. In ancient times only some tens of coins were usually struck at acceptable quality from the same die. The die had then to be engraved anew. Therefore the data of the ‘cultural database’ to which a particular query is submitted is inherently uncertain or approximately known.

Ideally our task is to search a sample q within a database containing perfect representations of the corresponding item r (see Fig.8 below). Item q is an imperfect / incomplete image of r . Therefore the outcome of this search can be four fold: (i) *correct identification* of an existing r matching with q , (ii) *correct negative answer*, i.e. q cannot be matched to any data base item r , (iii) *false alarm*, i.e. erroneous matching of q with some r and (iv) *missed detection*, i.e. failure to identify an existing r matching with q . Outcomes (iii) and (iv) are sometimes defined slightly differently under the terms ‘recall’ and ‘precision’. Let us assume that *false alarm* and *missed detection* occur with probabilities f_q , respectively m_q .

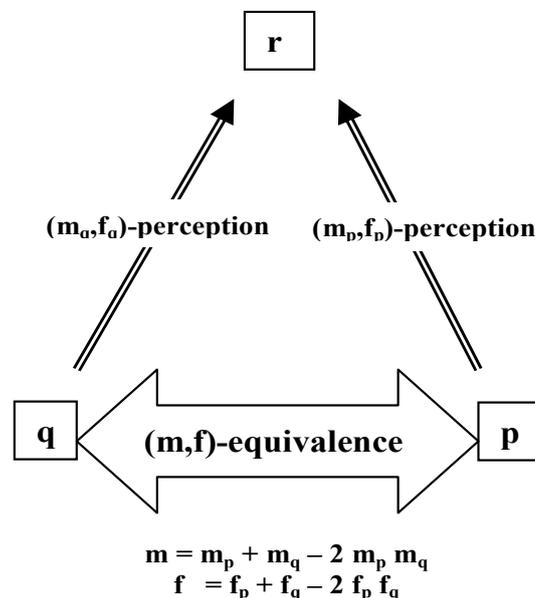


Fig.8. Equivalent Perception

In reality though, r is inaccessible (the lost canceller of the cancellation or the ‘ideal’ die of the engraver) and q can only be matched with a fictitious p being itself an imperfect image of r . The ‘cultural database’ consists of all p ’s deduced or

reconstructed from all existing samples to the best of our knowledge. The relationship of p to r is also characterized through probabilities f_p and m_p in the same way as before. Since comparing q to r is impossible, we have to compare q to p , see again Fig.8. It can be deduced that matching q to p occurs with probabilities f and m as expressed in Fig.8. These entail approximately the sum of the individual uncertainties m_p and m_q , respectively f_p and f_q . We are forced to an *equivalent perception* as depicted in Fig.8. There cannot be a straightforward search in an absolutely correct data base r , but only to an approximate proxy p . However under the allowances for the formulae for f and m , this can be seen as equivalent.

Equivalent perception in the textual content can be quite sophisticated and domain knowledge intensive. Careful trade-offs have to be drawn between m (missed detections) and f (false alarms). In the case of OCIT there are no issues of quantitative (efficiency in string matching) but only of qualitative nature. Even here, general approximate string matching approaches ([10]) are not applicable. Blind algorithmic and automated solutions are of no much help, if not enhanced with detailed domain knowledge. Character combinations in several languages and scripts have to be represented in all possible renderings, taking into account possible simplifications used by the engraver of the canceller or die, common pronunciation and spelling errors etc. All possible uses, misuses and omission of the different diacritics have to be foreseen. The initial (rightmost) khah in hane خانہ (see Fig.2) is indeed a khah (خ), but also with almost equal frequency a hah (ح). So خانه (khane) has always to be interpreted also as حانه (hane) and vice versa. The difference is only a not-so-easy to identify dot and a 0.5 missed detection probability would occur if a case like this is not meticulously foreseen. On the other hand خ and ح cannot be indiscriminately interchanged everywhere. This would blow up the false alarm probability f . A commonly agreed approach, methodology and collection of concrete equivalences for different languages and their versions across centuries would be highly desirable.

Textual content comparisons in the context described above, partly falls into the provisions of Recommendation 1 par. c of the Chicago Statement [7]. Editions or excerpts of the same work are identifiable as the instances p of our model. These are all imperfect images of r , the lost original manuscript of the author. We differ

however again in quantitative and qualitative terms. A large number of instances p is desirable and possible in our cases (e.g. cancellations, coins). The analysis involved in our comparisons does not address more than simple textual content of place names in small phrases and expressions.

Distributed Deployment

Nowadays, a tool like OCIT can be developed for a distributed deployment almost just as easily as in its present form. The problem lies in the willingness of interested parties to adopt, coordinate and sustain such an operation. There are several degrees of distribution. A ‘centralised’ one, around a server in the conventional sense and a ‘truly distributed’ pier to pier one, where all players have equal roles and responsibilities not only in the operation but also in the evolution (see below) of the environment. A ‘centralised’ operation is technically straight forward, but carries the difficulty of sustainable human involvement in a case with no apparent material rewards. The ‘truly distributed’ operation can draw much more resources from voluntary work and the enthusiasm of hobbyists, but relies on substantial technical challenges. Distributed updates and various degrees of collaboration are required. At the purely operational level, solutions exist for the operation proper. The following paragraphs investigate issues toward this goal and examine ways for a jointly administered evolution of such a distributed application targeting cultural items. As always in this work these are supposed to be ‘strikes’ of inaccessible ‘dies’.

Schema Driven Application

The storage, presentation and simple manipulation of a data item representing a cancellation (or a coin) can become truly generic. After all only CRUDE (Create, Update, Delete) actions are involved accompanied by simple logic. The main functionality concerns whatever searching possibilities in a relational data base could be generically described in a formal way. The parameterisation of the latter can be embedded into a corresponding XSD (XML Schema Definition Document). Hence a wide set of interested users can agree to a common functionality, entirely embedded and driven by an agreed schema. This functionality covers the presentation, storage

and search of the data inside a set of equally structured items. The scheme follows the diagram of Fig.9. The scenario shown is a three tier setup, whereby the user maintains a server and database and views/offers for viewing his collection via http. The underlying environment could however be even simpler, i.e. a Windows PC with local viewing via forms, e.g. OCIT. The ‘Logic’, ‘dbAccess’ and presentation (via GUI elements, possibly aided by embedded code, e.g. Javascript) are fully generic components (e.g. dll’s and form or html controls) consulting an XSD. The latter, not only imposes the data item’s structure, but also determines the way of its handling, in particular the parameters and structure of conceivable related queries. Notice that the community of users is not required to operate the same environment, but only ‘*Some Framework*’ allowing the porting of the generic components. Heavy server based players (e.g. a museum) and common users can then exchange, store and manipulate data item XML’s. These exchanges are not shown in the figure. Notice that the ‘museum’ and the ‘community of users’ around it cooperate on a purely pier to pier basis. They (i) can store the same or different items within the same family as defined by the XSD, (ii) have the same opportunities and predefined queries for searching such items either locally or remotely, (iii) can exchange, view and offer to viewing those contained in their own data base repository. This maintains a community of equals irrespective of size, equipment or daily effort invested in the field.

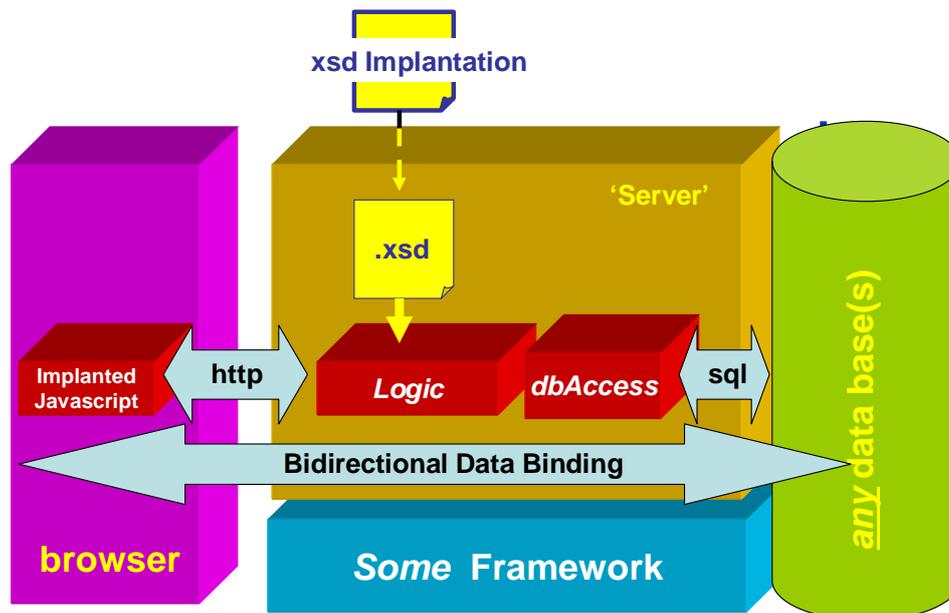


Fig.9. Schema Driven Implementation for Handling Data Items

Embedding the definition of all handling actions into a document like the XSD, allows a number of community wide cooperation and evolution paths. Upon agreement, another XSD brings new (hopefully upgraded, expanded) functionality. There is no need for any change, downloading of code or user intervention requiring special skills. The only problem lies with the data of items already stored into the data base. We now turn attention to this point. Conformant to our setup, we henceforth restrict our discussion to scenarios concerning the evolution of the XSD itself.

Collective Evolution and Schema Homogenisation

Suppose now that in the course of the collective use of an XSD corresponding to a collection activity by a community of users, some upgrades are planned. One possibility is a true superset to the current XSD, however other more complicated relationships to the original XSD are possible, see Fig.10. A concrete OCIT driven example is as follows. Suppose some user(s) decide to collect, scan and include in the data base postcards with late 19th – early 20th century images of the actual post office

sites or buildings. When distributed, the new XSD will provoke a new, updated data base schema as well as new viewing controls, probably also an entire new web page or form. This concludes the *structural update*, however a crucial problem remains: the population of the new data schema with the content of the original repository.

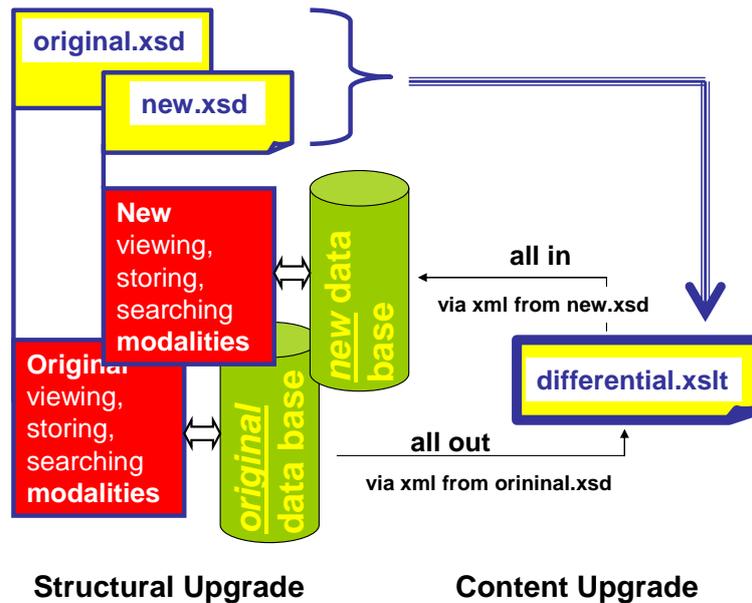


Fig.10. Structural and Content Upgrade

Here XSLT (XML Stylesheet Transformation Language) technology can provide the solution. In the same XML technology based spirit, the new XSD should be accompanied by an XSLT document capturing the difference from the original to the new XSD. Such an XSLT document caters for the mapping of XML documents validated against the *original XSD* into XML documents validated against the *new XSD* and featuring any amount of detailed structural modifications. To populate the new content base, the user only needs, on an item by item basis, to (a) read the data from the original data base and export it in xml form, (b) pass this XML through the XSL Transformation, (c) write the transformed XML into the new data base. Steps (a) and (b) constitute nothing new since these are already provided by the general set up of the previous paragraph (Fig.9). Step (b) can be a local capability or can be offered as a service. In either case generality is preserved throughout, with all *content*

upgrade functionality entailed now in the XSLT. Evidently content upgrade as described leaves the new structure with empty/default entries for new items (lost entries for those not envisaged in the new XSD). A further useful functionality would be the automated prompting or flagging in order to inform the user about the new, by now established schema. It is then up to him to care for the inclusion of available material (in our case scans of post office postcards) into the new, upgraded structure.

Aggregation under common hierarchy

Items of two *or more* same level collections can be easily aggregated under a new expanded hierarchy. The case is largely a derivative of the development presented in the previous paragraph. It has however some salient interesting characteristic, in particular the involvement of *more than two* XSD's. Let us draw an example from numismatics.

We consider an activity like the collection of 'Hellenistic Kingdoms' coins to which a particular subgroup is interested. At some point in time a dynamic modification / expansion would be desirable for serving other *same level* groups as well, e.g. to include 'Dynastic' issues, or perhaps an expansion uniformly across all coins representing 'Humans and Deities'. Under joint agreement all existing entries should then be map able to the new expanded structure. This mapping would in simple cases represent the union of all features of the individual subgroups. Or, it might constitute a more sophisticated object oriented paradigm under which representation of 'Humans and Deities' would acquire a parent schema role. Representation of 'Olympic Deities', 'Hellenistic Kings' and 'Dynasts', would then follow schemas derived from the 'Humans and Deities' parent. The challenge here lies not in an a priori design of these relationships, but in an evolutionary and collaborative derivation of these through simple ad hoc established practices. An aggregating template of an XSLT document draws in this case the particular XSD's (Hellenistic Kingdoms, Dynasts, Olympic Deities) and places these under a parent aggregation layer dealing with 'Humans and Deities'. Nothing prevents this broad structural expansion to be combined with detailed additional modifications across the old hierarchical levels. For instance 'named entity identification' either as stand alone services or as globally

accessed knowledge bases are foreseen in the Chicago Statement [7], Recommendation 1 par. d. What we postulate here in the incorporation of such links to a dynamic and distributed cataloguing process, with retrospective imposition of collectively defined and evolving schemata.

It is clear that aspirations as the ones outlined can only be based technically in the context of XML / XSD / XSLT. Such a scenario is depicted in the following Fig. 11 and constitutes an entirely off line upgrade procedure. Admittedly this also represents a process where some recognised authority should take the lead and responsibility in an otherwise pier to pier scheme. Maintenance and control of the XSDs should be also centrally administered. Otherwise a proliferation of schemata would quickly ruin the whole endeavour.

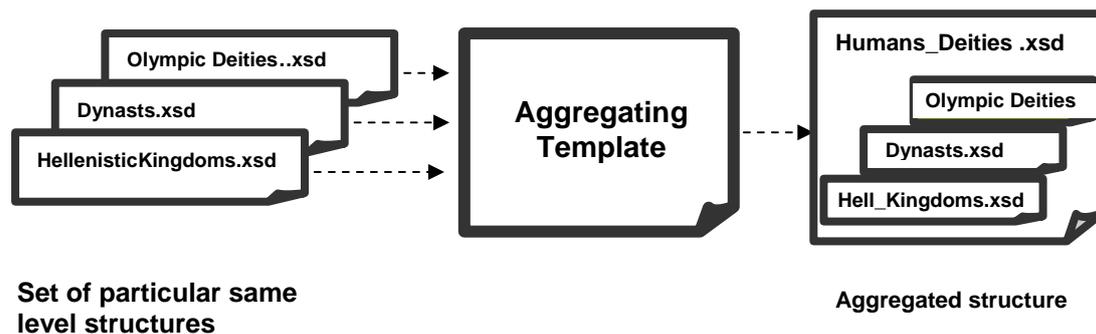


Fig.11. Dynamic and Distributed Schema Evolution & Aggregation

As before, the completion of the above scenario involves two phases. A *'design phase'* would comprise the generation of the schema hierarchy through involvement of the key players in each particular subfield. This might include a jointly agreed trial phase where the new schemata are tested in the field. This means entering, updating and searching a limited number of instance data in the operational distributed environment. Supposing this trial phase converges to a general agreement, a second *'deployment phase'* follows. New item presentation forms should be automatically generated prompting the user to enter the new additional data under the new schema hierarchy, upon visiting any 'old' item.

Numerous other aggregation patterns are conceivable and references in the spirit of [12] contain not only valuable ideas, but also ready to apply recipes in the form of XSLT templates.

CONCLUSION

Efforts like the CCO Project Development ([9]) are providing the groundwork of agreement on representation format and metadata of individual cultural objects. In a slightly different setting, we have addressed a framework of cataloguing one-of-a-kind objects, which are known and searchable only through (possibly a large number of) imperfect images thereof. The deployment of a tool like OCIT to collectors, i.e. to a large body of keenly interested individuals should allow a collective expansion of a cultural data base with ever new findings and features. A *quantitative expansion* of the content amounts to a greater number of entries. It presents no technical difficulty other than provisions for authentication and rights related to profiles and roles of users. However a *dynamic and distributed schema evolution* is extremely more challenging and interesting.

In the latter part of this work we have considered a widely distributed environment, not demonstrable in the present form of OCIT. This targets a community of users particularly interested in such a field. Collectors like philatelists might want to share their collection in a virtual (never real!) setting. In other cases museums, as larger but still pier to pier players, might want to join in collaborating toward a quantitative and preferably qualitative upgrade of cataloguing and searching activities. The use of generic components used as common denominator can shift all relevant requirements in the area of XML technologies, i.e. in the formulation and exchange of XSD and associated XSLT documents. This opens the way for a distributed and collaborative environment, where simple users can be part of quite elaborate mechanisms without ‘getting dirty’ with technology. A salient feature of an environment as presented is the possibility of gradual build up by enriching the structure and interrelationships of represented items. Moreover the possibility of aggregating ‘island communities’ opens up another important way for the digital preservation of cultural items through

the widest possible involvement of interested institutions and individuals. Future work is planned according to the conclusions just drawn: a pilot OCIT-like environment incorporating the basic technological choices and in parallel awareness creation and demonstration activities to encourage the adoption of the methodology to other areas of interest.

LITERATURE

- [1] Coles J.H. and Walker H.E., (1992), *Postal Cancellations of the Ottoman Empire* (in four volumes), Christie's-Robson Lowe, London.
- [2] Brandt O. and Ceylân S., (1963), *Türk Postaları İlk Filatelik Damga ve Mühürleri 1863 – 1920 - Premières Marques Postales Philateliques de la Turquie*, Pulhan Matbaası, İstanbul.
- [3] Nuhoğlu H.Y. and Mert T., (1990), *PTT Müzesi Osmanlı Posta Damgaları Katalogu*, IRCICA, İstanbul.
- [4] Nicolas A. and Galinos A., (1996), *Ξένα Ταχυδρομικά Γραφεία και τὰ Σήμαντρά τους στὰ Ἑλληνικὰ Ἐδάφη – Foreign Post Offices and their Cancellations in the Helladic Territories*, Collectio, Athens.
- [5] Birken A., (1992), *Philatelic Atlas of the Ottoman Empire*, The Author, Hamburg.
- [6] Mitchell T.F., (1953), *Writing Arabic, A Practical Introduction to the Ruq'ah Script*, Oxford University Press, New York.
- [7] BLACKWELL C. et al, (2008), *Classics in the Million Book Library*. Available from <http://www.stoa.org/million/chicagostatement.pdf> ; accessed 16 May 2008.
- [8] Anderson C., (2006), *The Long Tail: Why the Future of Business is Selling Less of More*, Hyperion.
- [9] CCO, (2006), *Cataloguing Cultural Objects: A Guide to Describing Cultural Works*. Summary available from <http://www.vraweb.org/ccoweb/cco/index.html>; accessed 17 May 2008.
- [10] Graham A. S., (1994), *String Searching Algorithms*, World Scientific.
- [11] THE J. PAUL GETTY TRUST, (2007), *Getty Thesaurus of Geographic Names* ® Online. Available from http://www.getty.edu/research/conducting_research/vocabularies/guidelines/tgn_1_cointents_intro.html#1_1_3; accessed 15 May 2008.

[12] Mangano S., (2005), XSLT Cookbook, O'Reilly.