# Annotations as a Tool for Disclosing Hidden Relationships Between Illuminated Manuscripts

Maristella Agosti, Nicola Ferro, and Nicola Orio

Department of Information Engineering, University of Padua, Italy
{agosti, ferro, orio}@dei.unipd.it

**Abstract.** Image digital archives of illuminated manuscripts can become a useful tool for researchers in different disciplines. To this aim, it is proposed to provide them with tools for annotating images to disclose hidden relationships between illustrations belonging to different works. Relationships can be modeled as typed links, which induce an hypertext over the archive. In this paper we present a formal model for annotations, which is the basis to build methods for automatically processing existing relationships among link types and exploiting the properties of the graph which models the hypertext.

## 1 Introduction

The ideas and concepts reported in this paper build upon our experience on the analysis of the user requirements, the design of a methodology, the development of a prototype system, and the analysis of the feedback from real users of a digital archive of historical material. The archive aims at the study and research on *illuminated manuscripts*, which are books, usually handwritten, that include illustrations and, in the past centuries, were manually and artistically decorated. Illuminated manuscripts are still the subject of scientific research in different areas, namely history of arts and history of science, and all the disciplines that are related to the subject of the book – e.g., botany, astronomy, medicine. Before the invention of photography, illuminated manuscripts have been the main mean for the dissemination of the scientific culture, and to this end play a major role as witness of the cultural heritage of different cultures, in Europe, Asia, and in the countries under the influence of the Arabic culture. The particular application to the cultural heritage domain poses interesting problems and challenges as reported in [5].

According to reached results [1,2] and a deep study about annotations [3,4], the use of annotations has been proposed as a useful way of accessing a digital archive, sharing knowledge in a collaborative environment of researchers, disseminating research results to students, automatically analysing user's annotations, with the aim of highlighting inconsistencies and suggesting new relationships among the images of the digital archive. The use of annotations as a research tool in the humanities has been reported also in [6].

The prototype digital archive of illuminated manuscripts that has been developed within our research activities has been called IPSA (Imaginum Patavinae

R. Basili and M.T. Pazienza (Eds.): AI*IA 2007, LNAI 4733, pp. 662–673, 2007.

Scientiae Archivum, archive of images of the Paduan science) [1,2], because the main focus of our initial project was to provide a tool for the analysis of the role of the Paduan school during the Middle Ages and the Renaissance for the spread of the new scientific method in difference sciences, from medicine to astronomy. IPSA is a case study for our research on methodologies and tools for researchers and scholars working on the study, the preservation and the dissemination of the cultural heritage. In this paper we focus on the modelling of additional functionalities that will be developed in the next release of the prototype.

The paper is structured as follows. In Section 2 we introduce the motivations and objectives of our work. The formal model for annotations of the content of a digital archive is presented in Section 3, and the methods exploited for automatic suggestions are presented in Section 4. Conclusions are drawn in Section 5.

## 2 Motivations and Objectives

Even if the primary goal of images in illuminated manuscripts was to represent the reality, during the Middle Ages authors of drawings were more interested on aesthetics than on realism and the primary role of illuminated manuscripts as a tool for scientists was lost. Images were often copied from or inspired by existing manuscripts, while the resulting drawings became increasingly different from the subjects they should represent. For researchers, it is of primary importance to state if drawings of a illuminated manuscripts are copied from previous manuscripts or if they are directly inspired by the nature. The disclosure of a link between two images belonging to two independent manuscripts, because one was the source for creating the other, allows to draw connections between the art of natural representations through the years and across the countries.

This is one of the main reasons why illuminated manuscripts are still the subject of research, for which a digital archive such as IPSA has to provide support for a number of particular users' needs.

In the following we discuss the major outcomes highlighted by the user requirements.

### 2.1 Disclosure of Relationships Among Manuscripts

As mentioned, it is of primary importance for researchers to discover if illustrations have been copied from images of other manuscripts, if they have been merely inspired by previous works, or if they are directly inspired by nature. A major requirement thus regards the possibility of enriching the digital archive by highlighting explicit relationships that have been discovered by a researcher. In particular, a research user should be able to create *links* that connect one image to another that it is related to, in some way. The analysis of user requirements on link management highlighted a number of advisable features that could be implemented.

- **Link authorship:** The creation of a link between two or more images depends on the scientific results of a researcher, who owns the intellectual rights

to the disclosure of a new relationship between images; for this reason the author of each new link has to be recorded by the system.
- **Link typology:** Since two images can be related for a number of different reasons, the kind of relationship should be explicit. Different typologies of links are envisaged to express the possibility that an image is the progenitor of a set of other images, or that two images are a copy of one another, and so on.
- **Paths:** Links may form *historical paths* among images, because images in a manuscript can be copies of another one which in turn are copies themselves of previous illustrations; hence two images may not be directly linked, because there is no direct relationship between them, but it could be possible to follow a path from one to the other by exploiting existing links.

It can be useful to clarify the notion of historical paths among images. A concept that has been introduced by researchers in the field of illuminated manuscripts is the one of *chains of derivation* among images. Each chain has a *progenitor*, which is an image that has been created through a direct examination of nature (i.e., a plant or a part of the human body). Other authors, who accessed the manuscript containing that image, may have directly *copied* or may have been simply *inspired* by that image. These new images may in turn be copied or be the source of inspiration of other authors and so on, creating a chain of references to previous works. Clearly, it may have happened that a same progenitor gave rise to more than one chain.

This requirement suggests the use of typed annotations that connects two manuscripts, two images, or even two parts of different images. These annotations, that have been called *linking annotations* need to have a type, which describes the kind of relationship between the two objects and provides a semantic to the link.

### 2.2 Dynamic Records and Intellectual Rights

Almost every digital archive dynamically changes over the years, mainly because of new acquisitions that increase the number of documents. This is also true for a digital archive of illuminated manuscripts, but there are other reasons that produce changes on the archive over time. The creation of records describing the documents and the images in a illuminated manuscript, as for any collection of historical works, is part of the scientific research itself. Some examples of changes to records are, for instance, that new relationships with other works have been discovered, or that the attribution to a given author became less certain.

Because creating a new record or modifying an existing one is part of the scientific work of researchers, the data management has to deal with intellectual rights. A researcher may prefer that some of the newly created records are not accessible by other users, at least until the results of his research have been published and his work have been acknowledged. This situation implies that users may decide which information can be shared with other users and which can not.

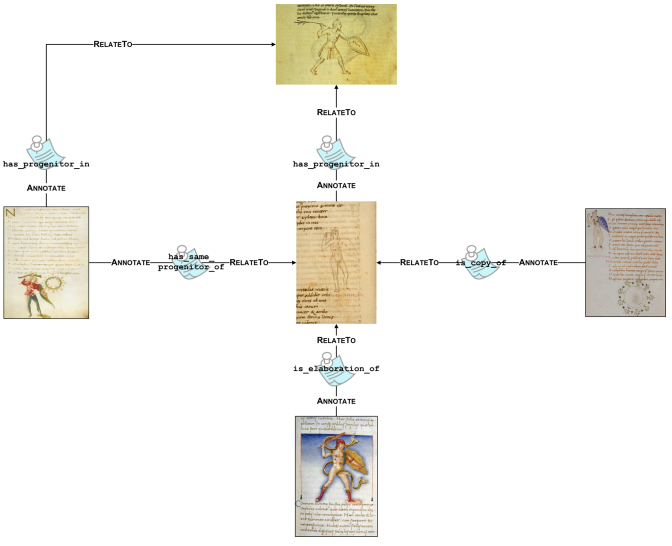


**Fig. 1.** Example of a personalized view on some linked images

This novel information, which is due to original results, should be stored in the digital archive at a different level than the information that is based on a general consensus. To this end, the use of annotations, both classical textual annotations and the proposed linking annotations, can be a viable tool providing that a user may state which annotations can be shared with the community or with his research group, and which ones has to remain private. Such a mechanism allows researchers for both using the digital archive as an advanced research tool and protecting their intellectual rights. Moreover, linking annotations add an hypertextual structure to the archive, which is different for each user and reflects his personal knowledge on the field, as shown in Figure 1.

### 2.3 Collaborative Environment

The study of illuminated manuscripts involves a number of researchers from different fields. In fact, illuminated manuscripts are of interest for both the historian of art and the historian of science, but at the same time, a herbal is of interest for the botanist because they represent plants and their possible variations through the centuries, a codex is useful for researchers on the evolution of civil and penal laws, an astrological book may give insights to researchers in medicine on the way stars where perceived to influence the health of people and to astronomers on how constellations where seen and represented. Hence, the scientific research on illuminated manuscripts involves a number of persons with different expertise, which should be able to cooperate in order to share their different knowledge and background.

As already mentioned, annotations that can be shared among researchers of different disciplines can provide an environment for collaboration and for the

sharing of knowledge and competencies. Furthermore, through the exploitation of linking annotations, each user may enrich the archive by a graph structure that reflects his knowledge and expertise. In general, graph structures may differ among researchers, because for instance a connection may appear evident for a botanist and not for a historian of arts. Sharing the set of connection, through tools for merging the different linking annotations and for analyzing the resulting graph, may ease the collaboration among researchers by pointing out inconsistencies or suggesting new possible relationships that have not been discovered yet.

## 3 A Formal Model of Annotations of Digital Content

As underlined in the previous section, the analysis of user requirements suggested that annotations, both in the form of text and in the form of typed links, can be a useful tool for a digital archive. This section presents the model of annotations that we are developing, aimed at a formal approach to annotations.

**Digital Object Sets.** An archive of illuminated manuscripts has to deal with different kinds of *Digital Objects* (DO). A preliminary user study highlighted that the objects that are studied by researchers are of three kinds: *manuscripts*, *pages* within a given manuscript, and *details* of pages, which usually are hand drawn images. We call them *Digital Contents* (DC), because they carry the information content that is the subject of scientific research.

The user study highlighted that a fourth DO has to be added to the digital archive: the *annotation* on digital content. Annotations are authored by researchers, and they may be either a tool for studying the collection of manuscripts – e.g., a way to highlight some interesting relationships that need to be further investigated – or the results itself of scientific research – the disclosure of new information about the DC in the archive. The following definition formalizes the different sets of DOs we need to deal with.

**Definition 1.** *Let us define the following sets:*

- $M$ *is a* ***set of manuscripts*** *and* $m \in M$ *is a generic manuscript.*
- $P$ *is a* ***set of pages*** *and* $p \in P$ *is a generic page. We define a function* $\mathrm{mp} : M \to 2^P$ *which maps a manuscript to the pages contained in it. The following constraints must be adhered to:* $\forall m \in M, \mathrm{mp}(m) \neq \varnothing$ *and* $\forall m_1, m_2 \in M, \mathrm{mp}(m_1) \cap \mathrm{mp}(m_2) = \varnothing$*, that is each manuscript must contain, at least, one page and pages cannot be shared among manuscripts.*
- $D$ *is a* ***set of details*** *and* $d \in D$ *is a generic detail. We define a function* $\mathrm{pd} : P \to 2^D$ *which maps a page to the details contained in it. The following constraint must be adhered to:* $\forall p_1, p_2 \in P, \mathrm{pd}(p_1) \cap \mathrm{pd}(p_2) = \varnothing$*, that is details cannot be shared among pages.*
- $DC = M \cup P \cup D$ *is a* ***set of digital contents*** *and* $dc \in DC$ *is a generic digital content.*
- $A$ *is a* ***set of annotations*** *and* $a \in A$ *is a generic* ***annotation****.*

– *$DO = DC \cup A$ is a **set of digital objects** and $do \in DO$ is a generic digital object.*

Note that *DO* (capital italic letters) is the set of defined digital objects, DO (capital letters) is the acronym for Digital Object and *do* (lowercase italic letters) is a digital object $do \in DO$. Similar considerations apply to digital contents and to annotations.

Each DO is uniquely identified by means of an *handle*.

**Definition 2.** *$H$ is a **set of handles** such that $|H| = |DO|$ and $h \in H$ is a generic handle. We define a bijective function $\mathrm{h} : H \rightarrow DO$ which maps a handle to the DO identified by it*[1]*: $\forall\, do \in DO, \exists!\, h \in H \mid \mathrm{h}(h) = do \Rightarrow \mathrm{h}^{-1}(do) = h$.*

*We will explicitly indicate when a handle identifies an annotation with the notation $h_a$, for the generic handle, and with $H_a \subseteq H$ for the subset of annotations handles.*

**Author and Group of Authors.** Each DO has an author who creates it. In the particular case of a digital archive of historical manuscripts, the author of a DC cannot interact with the present archive, because he lived centuries ago. On the other hand, nowadays researchers do not author manuscripts, pages, or details. For these reasons, we refer as *author* to only the users of the present archive that create annotations and that cannot create DCs. In our application scenario, groups of authors correspond to research groups, in which different researchers cooperate; a researcher may collaborate with different research groups.

**Definition 3.** *Let us define the following sets:*

– *$AU$ is a **set of authors** and $au \in AU$ is a generic author. We define a function $\mathrm{au} : AU \rightarrow 2^{H_a}$ which maps an author to the handles of the annotations authored by him.*
– *$GR \subseteq 2^{AU}$ is a **set of groups of authors** and $G \in GR$ is a generic group of authors. We define a function $\mathrm{gr} : AU \rightarrow 2^{GR}$ which maps an author to groups of authors he belongs to. The following constraint must be adhered to $\forall\, au \in AU, \mathrm{gr}(au) \neq \varnothing$, that is each author in $AU$ must belong to, at least, one group of authors.*

**Types of Annotation.** The type of annotation represents part of the semantics of an annotation. The following definition formalizes the notion of type of annotation.

**Definition 4.** *$T$ is a **set of types of annotation**, and $t \in T$ is a generic type of annotation.*

*The **types graph** is a labeled directed graph $(G_T, \mathrm{l}_T)$, where $G_T = (T, E_T \subseteq T \times T)$, $T$ set of vertices, $E_T$ set of edges, and $\mathrm{l}_T : E_T \rightarrow L_T$ with $L_T$ set of labels.*

[1] $\exists!$ is the *unique existential quantifier*, and it is read "there exists a unique ... such that ...".

The goal of the types graph is to provide some sort of structure and hierarchy among the types of annotation in order to navigate and browse through them.

As we discussed in the previous section, we need the possibility to express a relationship between DCs in the archive through the use of a *linking annotation* and the definition of the type gives us such possibility. Linking annotations are divided in two groups, that is the types graph can be partitioned in two disjoint subgraphs, which express a *hierarchical* or a *relatedness* relationship respectively between DCs of the same set – $M$, $P$, or $D$.

**Scope of Annotation.** An annotation can have different scopes, i.e. it can be *private*, *shared*, or *public*.

**Definition 5.** *Let* $S = \{Private, Shared, Public\}$ *be a* ***set of scopes*** *and* $s \in S$ *is a scope. Let us define the following relations:*

- ***equality relation*** $=: \{(s, s) \in S \times S \mid s \in S\}$
- ***strict ordering relation*** $\prec$: $\{(Private, Shared), (Private, Public), (Shared, Public)\}$
- ***ordering relation*** $\preceq \{(s_1, s_2) \in S \times S \mid s_1 = s_2 \vee s_1 \prec s_2\}$

We assume that each annotation can have only one of the three scopes listed above. Note that $(S, \preceq)$ is a *totally ordered set*. The choice of three levels of scopes is motivated by the fact that, an annotation can either be: of general interest, that is the consolidated results of past scientific research (public); a tool for exchanging information on a research work carried out by a group of researchers (shared); a way to highlight an interesting aspect that needs further investigation before being submitted to other researchers (private).

**Annotation.** Now we can introduce a formal definition of annotation.

**Definition 6.** *An* ***annotation*** $a \in A$ *is a tuple:*

$$a = \left(h_a \in H_a, au_a \in AU, G_a \in 2^{GR}, s_a \in S, t_a \in T\right)$$

*where:*

- $h_a$ *is the unique handle of the annotation* $a$*, i.e.* $\mathrm{h}(h_a) = a$*;*
- $au_a$ *is the author of the annotation* $a$*, i.e.* $h_a \in \mathrm{au}(au_a)$*;*
- $G_a$ *are the groups of authors which can access the annotation, such that* $G_a \subseteq \mathrm{gr}(au_a)$*;*
- $s_a$ *is the scope of the annotation* $a$ *– Private, Shared, or Public;*
- $t_a$ *is the type of the annotation* $a$*.*

**Annotation-Based Hypertext.** Given that each type of annotations that is taken into account expresses a relationship between two DCs in the form of a typed link, we consider that existing DCs and user's annotations constitute a hypertext.

**Definition 7.** *The* annotation-based hypertext *is a labeled directed multigraph:*

$$\left(\mathcal{H} = (DC, A), \text{annotate}\right)$$

*where:*

- *DC is the set of vertices;*
- *A is the set of edges;*
- annotate : $A \rightarrow DC \times DC$ *is the edge-function, which puts an edge between two DCs* $dc_1$ *and* $dc_2$ *if and only if there is a relationship between them, which is expressed by the annotation a.*

*The following constraints must be adhered to:*

1. *a* $dc \in DC$ *cannot be put in relationship with itself, that is* $dc_1 \neq dc_2$*;*
2. *the two DCs connected by an annotation must be of the same type, that is* $dc_1, dc_2 \in M \vee dc_1, dc_2 \in P \vee dc_1, dc_2 \in D$*.*

The annotation-based hypertext is built by putting an edge between two DCs vertices, if an annotation between that two DCs exists. Note that edges can be put only between DCs and not between annotations: this means that an annotation cannot connect other annotations. The two constraints on the annotation-based hypertext are based on a study carried out on the user requirements of the researchers that will access and annotate the digital archive: annotations do not have to express a relationship of a DC with itself, or with DCs of different kind. Since there are no constraints on the number of annotations that connect a pair of vertices, we deal with a multigraph. The existence of multiple edges between the same pair of vertices allows us to express different kinds of relationships between two DCs. In this way, we take into account both the possibility of different interpretations of the same contents given by independent authors, and the partial results of a same author, who is studying a particular subset of the digital content and expresses alternative relationships that need further investigations. Users are not expected to access the whole annotation-based hypertext, because annotations have scopes that are related to user's access rights. Thus, the following definition introduces an operator suitable for choosing the subset of the annotation-based hypertext that can be accessed by a user.

**Definition 8.** *Given an annotation-based hypertext* $\mathcal{H}$*, we introduce a* ***projection operator*** *that can have the forms:*

- $\mathcal{H}^\pi = \pi\left(\mathcal{H}, AU_\pi, S_\pi, T_\pi\right)$*, with* $AU_\pi \subseteq AU$*,* $S_\pi \subseteq S$*,* $T_\pi \subseteq T$*, constructs a new annotation-based hypertext* $\mathcal{H}^\pi \subseteq \mathcal{H}$ *such that:*

$$\begin{cases} A^\pi = \{a \in A \mid au_a \in AU_\pi \wedge s_a \in S_\pi \wedge t_a \in T_\pi\} \\ DC^\pi = DC \end{cases}$$

- $\mathcal{H}^\pi = \pi\left(\mathcal{H}, GR_\pi, S_\pi, T_\pi\right)$*, with* $GR_\pi \subseteq GR$*,* $S_\pi \subseteq S$*,* $T_\pi \subseteq T$*, constructs a new annotation-based hypertext* $\mathcal{H}^\pi \subseteq \mathcal{H}$ *such that:*

$$\begin{cases} A^\pi = \{a \in A \mid G_a \in GR_\pi \wedge s_a \in S_\pi \wedge t_a \in T_\pi\} \\ DC^\pi = DC \end{cases}$$

*Both operators have a generalized version, where the $\star$ symbol can be replaced to an input parameter in order to express that the whole set has to be used.*

*For example* $\pi(\mathcal{H}, AU_\pi, S_\pi, \star) = \pi(\mathcal{H}, AU_\pi, S_\pi, T)$.

This operator provides us with a personalized view for the user of the annotation-based hypertext $\mathcal{H}$. The first form allows us to select edges on the basis of author(s), scope(s) and type(s) of the annotation, while the second form utilizes groups of authors instead of authors as selection criterion. This operator is quite flexible, if combined with the previous definitions. For example, if, given an author $au \in AU$, we want to extract the subgraph with all the public annotations (edges) inserted by authors that belong to the same groups of $au$, we can use $\mathcal{H}^\pi = \pi(\mathcal{H}, \mathrm{gr}(au), \mathrm{Public}, \star)$. Finally, the expressive power of this operator can be further enriched by using also the usual union, intersection, and difference set operators. The projection operator represents the standard way for a user to perceive the annotation-based hypertext, because a user is not allowed to access all the edges of the hypertext but he can access only the public ones, those belonging to him, and the ones shared with groups of authors the user belongs to.

**Definition 9.** *Let us define the* ***annotation compatibility set*** $C \subseteq A \times A \times [0,1]$ *that expresses the degree of compatibility of the types of annotation among given pairs of annotations, where* 0 *means no compatibility at all and* 1 *means full compatibility. Let us define the* ***compatibility score*** $\mathrm{c}(C, a_1, a_2) = c \in [0,1]$ *between two annotations given an annotation compatibility set* $C$*, which returns the compatibility* $c$ *between two annotations if* $\exists\,(a_1, a_2, c) \in C$.

The actual value that $c$ assumes for different pairs of annotation types is part of previous knowledge about the semantic of the annotation types and on their organization in the types graph. We assume this value is given by specialists in the field of illuminated manuscripts.

**Definition 10.** *Given an annotation-based hypertext* $\mathcal{H}$*, a set* $T$ *of annotation types, and a types graph* $G_T$*, we introduce a* ***pair-wise compatibility operator*** $\xi(\mathcal{H}, T, G_T, dc_1, dc_2) = C_\xi$ *that* $\forall a_1, a_2 \in A \mid \mathrm{annotate}(a_1) = (dc_1, dc_2) = \mathrm{annotate}(a_2)$ *returns a compatibility score for the annotations connecting* $dc_1$ *and* $dc_2$.

*Given an annotation-based hypertext* $\mathcal{H}$*, a set* $T$ *of annotation types, and a types graph* $G_T$*, we introduce a* ***path-wise compatibility operator*** $\xi(\mathcal{H}, T, G_T, dc_1, dc_2, dc_3) = C_\xi$ *that* $\forall a_1, a_2 \in A \mid \mathrm{annotate}(a_1) = (dc_1, dc_2) \wedge \mathrm{annotate}(a_2) = (dc_2, dc_3)$ *returns a compatibility score for the annotations connecting* $dc_1$ *and* $dc_2$ *with respect to the annotations connecting* $dc_2$ *and* $dc_3$.

Both forms of the compatibility operator make use of the types graph, which expresses the relationships among the different types of annotation, in order to determine the degree at which the type of two different annotations is compatible. Note that the annotation compatibility set $C$ can be used to produce a ranking among the annotations connecting different DCs in order of severity of compatibility problems.

## 4 Automatic Suggestions of Relationships Among DCs

The introduced model and operators can be exploited to create tools for helping the user of a digital archive to perform scientific research on its content. In particular, the annotation-based hypertext can be automatically analyzed to highlight possible inconsistencies among the annotations – e.g., two DCs are annotated with typed annotations that have different and possibly contrasting semantics – as well as to extract new information about possible relationships – e.g., two DCs are not annotated but the surrounding set of edges suggest the possibility of a relationship among them. It has to be stressed that the automatic analysis of the graph can only provide the user with *suggestions* on possible new or different annotations. The final choice of which annotations are to be added or modified is made by the research user who, from his cultural and scientific background, can take the final decision on relationships among digital content. Moreover, the research on illuminated manuscripts is an ongoing work, for which temporary inconsistency and incompleteness are normal events. Yet, the automatic analysis may help the researcher by suggesting the creation of new annotations, because the task of accepting to author an automatic annotation is expected to be simpler than creating an annotation from scratch.

### 4.1 Suggestions of Possible Inconsistencies

As previously explained, the model allows for multiple annotations of the same pair of digital contents. This means that public annotations of different authors may be different, or that for a given author, public annotations may differ from private ones, or even that there can be different private annotations. These inconsistencies may be made on purpose, but may also be the result of an erroneous interaction with the system, or to changes in the view of the DC relationships over the years. In any case, the analysis of the graph may pinpoint particular relationships that need to be carefully checked by the user.

**Definition 11.** *Given an annotation-based hypertext $\mathcal{H}$, a set $T$ of annotation types, a types graph $G_T$, a subset of authors $AU_\psi \subseteq AU$ and a subset of scopes $S_\psi \subseteq S$, the* ***pair-wise inconsistency finder operator*** *$\psi(\mathcal{H}, T, G_T, AU_\psi, S_\psi) = C_\psi$ firstly computes $\mathcal{H}^\pi = \pi(\mathcal{H}, AU_\psi, S_\psi, T)$ and secondly computes*

$$C_\psi = \bigcup_{\substack{dc_1, dc_2 \in DC^\pi \mid \exists a \in A^\pi, \\ \text{annotate}(a) = (dc_1, dc_2)}} \xi(\mathcal{H}^\pi, T, G_T, dc_1, dc_2)$$

*The* ***path-wise inconsistency finder operator*** *$\psi(\mathcal{H}, T, G_T, AU_\psi, S_\psi) = C_\psi$ firstly computes $\mathcal{H}^\pi = \pi(\mathcal{H}, AU_\psi, S_\psi, T)$ and secondly computes*

$$C_\psi = \bigcup_{\substack{dc_1, dc_2, dc_3 \in DC^\pi \mid \\ \exists a_1, a_2 \in A^\pi, \\ \text{annotate}(a_1) = (dc_1, dc_2) \wedge \\ \text{annotate}(a_2) = (dc_2, dc_3)}} \xi(\mathcal{H}^\pi, T, G_T, dc_1, dc_2, dc_3)$$

Once that either the pair-wise or the path-wise operator is applied, from the set $C_\psi$ it is possible to extract all the compatibility scores related to annotations made by authors $AU_\psi$ (possibly belonging to the same group) with scopes $S_\psi$. It is then possible to apply a threshold function on the set of compatibility scores, in order to provide the user with all the annotations that may be inconsistent or even contradictory. The degree by which two annotations are inconsistent depends on the semantics that the users give to the annotation types and to the types graph. The approach is general enough to support different definitions of the compatibility score, which are based on the knowledge of the application domain. In the case of illuminated manuscripts, the compatibility scores are based on the particular kind of relationships that can express a hierarchical relationship or a relatedness (and non-hierarchical) one. Examples of inconsistencies are that the same two DCs could be annotated as in hierarchical and non-hierarchical relationship at the same time, or that a $dc_1$ has been set as an ancestor of $dc_2$ by one author and viceversa for a different author. Suggestions of inconsistencies can be exploited in different way: as placeholders for highlighting unclear relationships between DCs that a user is interested in investigating in detail; as an indication of the more debated relationships in the archive.

### 4.2 Suggestions of New Relationships

The analysis of the annotation-based graph can highlight that two DCs are not annotated, yet there is a path that connects them. Moreover, since DCs are made of different sets which are organized hierarchically, the annotation of two manuscripts may suggest a similar annotation between two details, and viceversa. Also in this case, the existence of similar relationships may only suggest the presence of new relationships, which must be validated by the research user. Yet it can be considered that the presence of suggestions would ease the user in creating the network of annotations of the digital archive. Of course, there is also the possibility that the automatic analysis of the graph will disclose new relationships, at least for non expert users.

**Definition 12.** *Given an annotation-based hypertext $\mathcal{H}$, a set $T$ of annotation types, a types graph $G_T$, a subset of authors $AU_\psi \subseteq AU$ and a subset of scopes $S_\psi \subseteq S$ we introduce a* ***relationship finder operator*** $\rho\left(\mathcal{H}, T, G_T, AU_\rho, S_\rho\right) = C_\psi$ *that functions as follow:*

1. *compute* $\mathcal{H}^\pi = \pi(\mathcal{H}, AU_\rho, S_\rho, T)$
2. *compute the transitive closure* $\mathcal{H}^{\pi+}$ *of* $\mathcal{H}^\pi$
3. $\forall dc_1, dc_2 \in \mathcal{H}^{\pi+} \mid \nexists a \in A^\pi, \text{annotate}(a) = (dc_1, dc_2)$*, that is all of the DCs among which exists a path but are not directly connected, for each path* $P = dc_1 a_1 \ldots dc_m a_h dc_n \ldots a_k dc_2$ *connecting* $dc_1$ *to* $dc_2$*, compute* $C_{\rho,P} = \bigcup_{dc_{i_1}, dc_{i_2}, dc_{i_3} \in P} \xi\left(\mathcal{H}, T, G_T, dc_{i_1}, dc_{i_2}, dc_{i_3}\right)$
4. *if exists a path* $P$ *such that* $\sum_{a_1, a_2 \in C_{\rho,P}} \mathrm{c}(C_{\rho,P}, a_1, a_2) > T_\rho$ *(alternatively* $\prod_{a_1, a_2 \in C_{\rho,P}} \mathrm{c}(C_{\rho,P}, a_1, a_2) > T_\rho$*), with* $T_\rho$ *given threshold, than it suggests the existence of a possible relationship between* $dc_1$ *and* $dc_2$*.*

## 5 Conclusions

This paper describes an approach to the development of models and tools for a digital archive of illuminated manuscripts. We carried out an analysis of the user requirements for the use of the archive as a tool for scientific research. According to the user requirements, annotations have been suggested as the main functionality to be added to a digital archive.

We have proposed a formal model of annotations that introduces the notion of annotation-based hypertext and explores some of its properties, in order to automatically extract some relevant information about the relationships among digital contents and provide users with suggestions about possible inconsistencies between different results, and suggest the existence of new relationships.

## References

1. Agosti, M., Ferro, N., Orio, N.: Annotating Illuminated Manuscripts: an Effective Tool for Research and Education. In: JCDL 2005. Proc. 5th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 121–130. ACM Press, New York, USA (2005)
2. Agosti, M., Ferro, N., Orio, N.: Graph-based Automatic Suggestion of Relationships among Images of Illuminated Manuscripts. In: SAC 2006. Proc. 21st ACM Symposium on Applied Computing, pp. 1063–1067. ACM Press, New York, USA (2006)
3. Agosti, M., Bonfiglio-Dosio, G., Ferro, N.: A Historical and Contemporary Study on Annotations to Derive Key Features for Systems Design. International Journal on Digital Libraries (2007), http://dx.doi.org/10.1007/s00799-007-0010-0
4. Agosti, M., Ferro, N.: A Formal Model of Annotations of Digital Content. ACM Transactions on Information Systems (TOIS) 26, 1–55 (2008) (to appear)
5. Crane, G.: Cultural heritage digital libraries: Needs and components. In: Agosti, M., Thanos, C. (eds.) ECDL 2002. LNCS, vol. 2458, pp. 626–637. Springer, Heidelberg (2002)
6. Frommholz, I., Brocks, H., Thiel, U., Neuhold, E., Iannone, L., Semeraro, G., Berardi, M., Ceci, M.: Document-Centered Collaboration for Scholars in the Humanities – The COLLATE System. In: Koch, T., Sølvberg, I.T. (eds.) ECDL 2003. LNCS, vol. 2769, pp. 434–445. Springer, Heidelberg (2003)