DESIGN OF A NEW DATA STRUCTURE TO SUPPORT NON-INVASIVE DIAGNOSTIC ON HERITAGE METALS

Complete Research

Rosselet, Antoine, HES-SO // University of Applied Sciences Western Switzerland, HEG Arc, Neuchâtel, Switzerland, antoine.rosselet@he-arc.ch

Rochat, Vincent, HES-SO // University of Applied Sciences Western Switzerland, HEG Arc, Neuchâtel, Switzerland, vincent.rochat@he-arc.ch

Gaspoz, Cédric, HES-SO // University of Applied Sciences Western Switzerland, HEG Arc, Neuchâtel, Switzerland, cedric.gaspoz@he-arc.ch

Abstract

Conservation of heritage artifacts is a very sensitive task as conservators usually have very little information about the artifacts. Moreover, due to the uniqueness and the historic value of these artifacts, invasive analysis are not always possible. Therefore, without sampling options, conservators are required to use non-invasive diagnostic methods in order to identify the metal characteristics of the artifact. When confronted with an unknown artifact, conservators generate conceptual models of the corrosion forms. These models are based on formal representations of corrosion forms, but are not directly exploitable for drawing hypotheses regarding the underlying metal. This paper presents the design of a data structure generated from the conceptual models which supports the comparison and retrieval of corresponding artifacts. Integrated with a database of heritage artifacts, this data structure offers advanced decision support to conservators confronted with unknown artifacts.

Keywords: conceptual model, data structure, design science, decision support system

Introduction

The restoration of heritage artifacts requires careful identification of the material before any treatment can be performed. Over the years, the field of conservation-restoration of ancient or patrimonial objects evolved toward better preservation of the artifacts, mainly by reducing invasive analysis and applying very specific treatments in order to avoid further deterioration. These changes imply two assumptions. First, in order to reduce or totally abandon invasive analysis of the material, such as sampling, we need to be able to perform the same diagnostic with non-invasive techniques such as microscopic analysis. Second, in order to apply specific and targeted conservation treatments, restorers need to identify univocally the composition of the artifact.

In this paper, we present an innovative way of transforming non-invasive analysis of ancient metallic artifacts in actionable data structures that can be processed, compared and analyzed in order to predict the metallic composition of the artifacts. Currently, the most used non-invasive technique of heritage metals identification is stratigraphic analysis (Bertholon, 2000). This technique borrows from the field of geology and archeology by representing corrosion forms as a superposition of layers also called stratigraphies. After many years buried in soil or water various corrosion products and deposits (for example sediments) affect the metal contained in the artifact. The restorer visually observes these corrosion products and deposits in order to construct a representation of the different layers of deposits and corrosion products, like a geologist would do with a core sample.

Currently, there is no support tool to assist conservators in building their stratigraphies, mainly due to the novelty of the method. It is mainly a pencil/paper method, relying on defined construction rules (Figure 1). In order to support the broader diffusion of the method, its initiators are looking for a tool to assist conservators in using it. However, even if the term is not used by the authors of the method, the analysis of corrosion products using stratigraphies is some sort of conceptual model of the analysis of historic corroded metals. However, we can then build on the rich experience of conceptual modeling and data structures to design a tool to assist restorers in identifying underlying materials in conservation projects on historic artifacts.

In this paper, we start by introducing the notion of conceptual modeling and show how it is related to the stratigraphic methods developed by Bertholon (2000). Then, we draw on the literature on data structure to show how we use the generated conceptual model to produce a digital model (data structure) of the corrosion products of the artifact. We then explain how we designed the digital model and how we translated the conceptual model into a graph model that can be used to analyze and compare stratigraphic representations from various artifacts. Finally, we present some scenarios in which the data structure can also be used to increase its value for professionals in the conservation-restoration field.

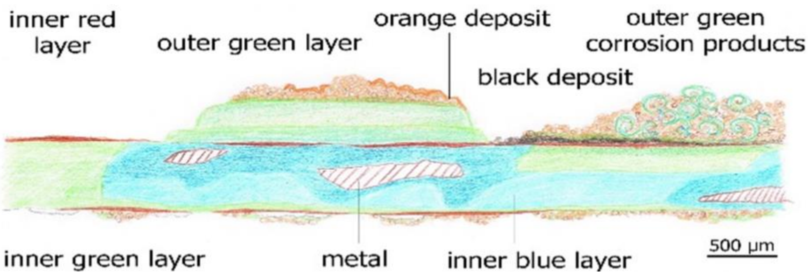

Figure 1. A stratigraphy, as drawn by a conservator

The principal contribution of this paper is a new data structure that can be used to store stratigraphies of heritage artifacts in a graph database in order to support professionals in identifying the metallic composition of these artifacts without invasive sampling. The subsidiary contribution is a design process that can be used to tackle similar problems that need to generate digital representations of analogue models in order to process them in some way.

From reality to data structure

Conceptual models are widely used in the information systems (IS) field (Wand & Weber, 2002). Kung and Solvberg identified their main purposes: (1) supporting communication between developers and users, (2) helping analysts understand a domain, (3) providing input for the design process and (4) documenting the original requirements for future reference (as cited in Wand & Weber, 2002). In management science, models are defined as “a representation of reality intended for some definite purpose” (Pidd, 2010). These definitions contain two elements that are related to our research objective. First, because we are not able to sample the observed artifacts, we will not be able, during the creation of the conceptual model, to univocally identify any parts of the reality. So, we will have to rely on observations of the artifacts in order to create a stratigraphy representing its corrosion products. Second, the model is created to achieve a specific goal. We are not trying to have a model representing the shape, color, texture or hardness of the object. Rather, we are seeking to create a model that will contain sufficient information to help us determine the nature of the metal composing it. In order to support these goals, the conceptual model should abstract an appropriate simplification of the reality.

Furthermore, the conceptual model needs to be designed so that it can later be translated into data structures that are relevant to our goal and which will help us understand the problems we want to address. Indeed, many different conceptual models can be made from a same reality. Depending on our needs, some conceptual models are pertinent whereas others are not. This is why an analysis phase needs to be undertaken in order to design an appropriate and actionable conceptual model that can further be used (Wand, Monarchi, Parsons, & Woo, 1995).

One well-known model is the subway map that a public transport operator would produce. The map is a simplified representation of the physical network, which allows passengers to work out possible routings through the network of lines and stations. However, these maps deliberately distort the reality in order to support their purpose. Lines are distorted in order to emphasize their general direction and the interchanges, regardless of their physical layout. This partial representation of the reality fulfills its goal of supporting passenger routing, but is of no use for contractors dealing with network maintenance. Pidd has a more specific definition of the model that we will use in the rest of this paper. "A model is an external and explicit representation of part of reality as seen by the people who wish to use that model to understand, to change, to manage and to control that part of reality." (Pidd, 2010)

As we saw in the introduction, the analysis of corrosion forms on ancient artifacts requires specific knowledge of its constituent corrosion products. In order to be able to compare, analyze and discuss these corrosion products, we need to create a stratigraphic representation of them. According to Robinson (2010), the first model to create and which is based on the reality we want to understand is called the "conceptual model". It is the first step which can then lead to a data structure stored in a computer. Once stored, the data structure can be used to perform simulations and to find out behavior patterns from the real world. In our context, we want to use conceptual models (1) as a final product to have a visual representation of an artifact to work on and to talk about and (2) as an intermediate step linking reality to data structure.

To illustrate the first use, we can think of the business model canvas (Osterwalder, Pigneur, & Tucci, 2005). We start with the business we want to analyze (reality). From the structure and functioning of this business we can draw a canvas, which is a representation of the reality. It consists of a conceptual model and a common tool to convey and discuss ideas about the business. The model can also lead to the design of new business models. Another example is the design of a Business Process Diagram (BPD). Its elaboration starts with an observation of the real world—in this case how people are organized within an organization—and results in a diagram that follows the rules dictated by the Business Process Model Notation (BPMN) (White, 2004). People can then discuss, analyze and develop that model. As it stands, the model is sufficient to support its purpose without needing to translate it into a computer language.

Note that in the two examples above, the conceptual model generated followed the rules dictated by the canvas or the framework in which it operates. However, this is not necessarily the case. The main advantage of modeling within a framework is that people only need to know how the framework works to understand the model and therefore the real world it depicts. It also allows easy comparison of different models, as they use the same notation conventions.

The second use is what really interests us. Indeed, as we want to automate the process of stratigraphic representation, we need a computer system to help the user construct the stratigraphy and to store it in a database. A simple drawing cannot be directly used to run simulations or to perform further tasks. Therefore, we need a data structure representing the stratigraphy and supporting further processing (Peuquet & Marble, 1990).

From reality to conceptual model

In order to build an actionable data structure, we first need to understand how to translate the perceived reality in a conceptual model that can be shared between, and understood by, users. When designing electronic circuits, engineers start from a conceptual model showing the different components of the electronic circuit. Each component and its characteristics are represented by specific drawings in the model. This helps the engineer understand the circuit structure and to compare it with other circuits. This can also be used to compute variations in current along the circuit. The same analogy can be made in the context of workflow management. We start by modelling the main process supporting the workflow using common notations such as BPMN. Here again, the BPD is a conceptual model of a real process taking place in a company. One particularity of this model is that it does not look like reality. Indeed, we do not “recognize” a human activity when a rounded rectangle is drawn; it only looks like a rectangle, not like a human performing a task. The same is true for most of the BPMN components. However, someone who knows the notation conventions can easily understand the model and the real process beneath it. An effort of abstraction is needed, but once it is done, the process is understandable. Manual comparison between different BPDs is also made possible thanks to BPMN.

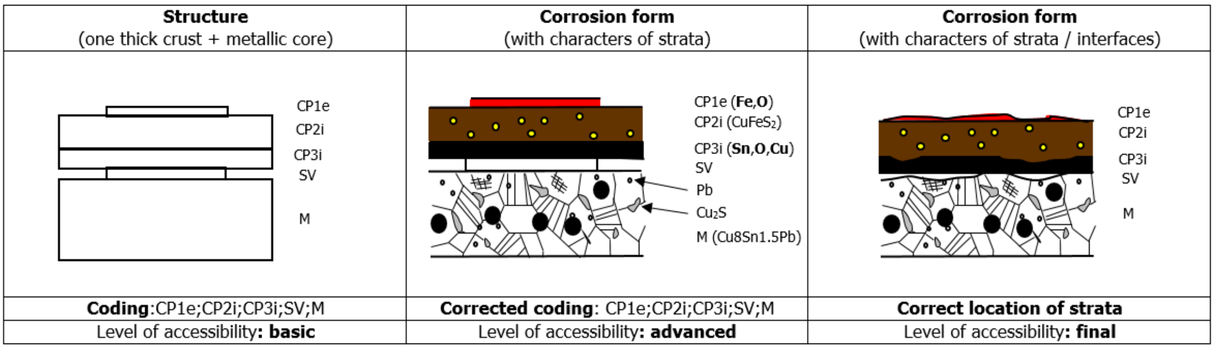

Figure 2. Conceptual models of stratigraphies

In our context, conservators will use specific notations to represent their visual observations. There is no wish to create a truthful representation of the artifact (like a map), but to create an abstraction of the reality in order to learn and share findings about the specific assembly of layers of corrosion products. Therefore, the conceptual model focuses on the area of interest for conservators: layers of corrosion products and their interfaces, as presented in Figure 2.

From conceptual model to data structure

Conceptual models are useful for sharing, studying and analyzing a specific view of the underlying reality. However, they are of limited use when computational processing is required in order to gather more information than the simple abstraction of the reality. Thus, if we want to apply computational processing to the conceptual model for further analysis, we need to transform it to a format that can be understandable to a computer. The translation is not always straightforward; refinements may be needed, depending on the conceptual model we use.

We have seen before that conceptual models can be used to represent the flow of current in a circuit. However, these models can only be used as a basis for understanding the functioning of the circuit. They are of little use in understanding what happens if a specific part of the circuit is altered. In order to comprehend the dynamics of the circuit, various software packages have been developed (Cellier & Clauss, Christoph, 2007) to support users in designing a circuit with building blocks—the blocks used in the conceptual model—and to derive equations from the model. These equations can then be used to simulate the model behavior, which helps deduce how the real electronic circuit works. The same applies to BPMN. In order to test the flow of the process we need a new language, the Business Process Execution Language (BPEL) (White, 2005), that can be executed in order to follow, step by step, the execution of the process. Therefore, BPEL gives us the ability to translate the conceptual model (a set of shapes) into an executable workflow.

However, many steps are needed in order to perform such a translation. When translating BPMN to BPEL, the BPD—made up of BPMN elements—first needs to be broken down into smaller parts. In this case, the graph structure is transformed into a block structure, with one entry point and one exit point for each BPD component. These parts are then individually translated into BPEL, depending on their type. For instance, a component that is identified as a sequence will be translated into BPEL with a sequence tag, whereas a component listed as a switch will display a switch tag. Each tag has its own characteristics, sub-characteristics and behavior. Therefore, business rules are taken into account, as a switch component does not behave like a while component and is composed of different elements. It is easier to respect these rules when the process is broken down into many components, as each one is individually analyzed and translated (Ouyang, Dumas, Aalst, Hofstede, & Mendling, 2009).

Corrosion product stratigraphies are made of superposed blocks representing a specific corrosion product layer. These layers can have characteristics such as color, hardness, friability, inclusions etc. Up to a point, these characteristics can be embedded in the representation of the strata. However, in order to retrieve similar stratigraphies, we need to be able to transform these graphical representations of corrosion products into an executable representation. A simple image analysis is not sufficient as we are interested in comparing the individual characteristics of the strata and not their appearance.

Contextually useful data structures

Finally, due to the fact that conceptual models are by definition simplifications of the reality, we need to find a way to add detailed information on parts of the models where information cannot be explicitly represented by the modelling conventions. For instance, in BPMN there is no notion of timing, or for how far the employees performing the tasks are from each other. In a map, we will generally not indicate the weather. Of course, some of these elements are not important. In fact, it depends on how we want to use the model; it depends on the data structure we need. As well as providing a common framework to work on and to talk about, the conceptual model allows us to imagine the visualization of the data structure; it becomes a prototype for creating computer solutions. When developers are asked to design software representing the data structure of a reality, they need to know the needs of the people who will use it and how they want to see their data—that is why it is so important in the process which leads to data structures.

In order to generate the appropriate data structures, it is essential to add some elements that we lost during the translation of the reality into a conceptual model. Indeed, as it is said for the BPMN: "the diagram itself will not display all the information required to create a valid BPEL file. A diagram with all that information would be too cluttered to be readable. A BPMN diagram is intended to display the basic structure and flow of activities and data within a business process. Therefore, a modeling tool is necessary to capture the additional information about the process that is necessary to create an executable BPEL file." (White, 2005)

The data structure needs to include this additional information in order to respect the business rules. White (2005) adds this information when generating a BPEL using a dedicated tool which knows what kind of extra elements are needed when translating BPMN to BPEL. Thus, the modeler has to give the right information to the tool so that the latter can create a correct BPEL file that can later be used for various applications. When the design of a BPEL is finished, it is then possible to analyze its structure and validate the underlying process after some modifications are applied. Schmidt and Stahl suggest transforming BPEL processes into Petri nets to do so. Therefore, it would be possible to verify properties, such as checking whether a customer will always get an answer in the analyzed process (Schmidt & Stahl, 2004). In order to generate data structures, we started from reality by going through conceptual data. Eventually, when we have an exploitable tool that can validate an existing process, the path has been reversed: based on the data structure, we can assess a real process. Thus, this assessment can be very useful from the perspective of a process enhancement or a deeper analysis.

On the one hand, it is essential to create a conceptual model from reality, so that we can come up with data structures based on this model. On the other hand, depending on the information we are looking for, it is important to add more elements to the data structure than the ones that are embedded in the conceptual model. This will provide us with data that can then be manipulated in order to perform further tasks, such as simulation. This process is called “supercharging”.

Non-invasive diagnostic on heritage metals

The conservation of ancient and historic metal artifacts requires a detailed understanding of both their composition and their alteration. Therefore, conservators have developed preventive conservation strategies and curative treatments in order to stabilize active corrosion processes, while preserving all relevant information. However, the invasive and/or destructive character of metallography and some associated chemical analysis is the main factor which limits their application to cultural heritage artifacts. In the past, the fields of archaeometallurgy, history of techniques and conservation science have produced large quantities of metallographic studies using samples and/or physical and chemical analyses (invasive, non-invasive, destructive and semi-destructive). These studies cover all pre-industrial metal families (Fe, Cu, Ag, Au, Pb and Sn). Unfortunately, this data rarely contains information on corrosion forms and the nature of the corrosion products (Degrigny & Senn, 2012).

For this project, the stratigraphic representations were modelled using the methodology developed by Bertholon (2000). It is based on the fact that the structure of artifacts can be broken down into several layers (strata), which together constitute a stratigraphy. A stratum can be the metal itself or any corrosion products that have affected the metal. Each of these layers has several characteristics depending on its nature. The goal of the conservator is to find out the metal of which the artifact is made. In order to do so, layers have to be removed and treatment has to be applied. Being able to compare the artifact with other artifacts that have similar corrosion products and which are listed and analyzed in a report is highly valuable for a conservator, as this could give clues about how to treat it.

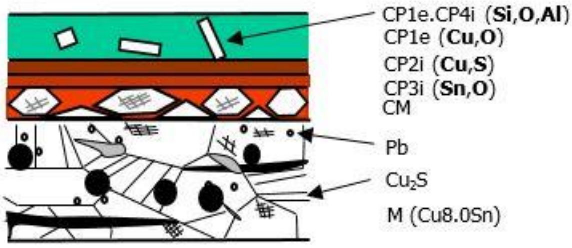

Figure 3. Example of a stratigraphy drawn on a computer

After different strata have been identified and manually drawn, they are reproduced in a computer, using special notation conventions as seen in Figure 3 (Bertholon, 2000). This constitutes the conceptual model. Then, the comparison is straightforward as all the artifacts drawn follow the same modelling framework. Nevertheless, the comparison is only visual; thus far, the stratigraphies have not been coded into a computer language that could compare them, based on objective characteristics.

Without a tool that can regroup the stratigraphies, the conservators have difficulty identifying the metal the artifact is composed of because comparing it with others is complicated. Therefore, applying any curative treatment is arduous. A remaining option would be to take an artifact sample. But as this is forbidden for most artifacts by professional ethics, the conservator has to find another way to identify the metal.

In order to apply the right restoration method to an ancient metal artifact without damaging it, the more information and comparison the conservator has at hand, the more effective the work will be. Therefore, providing the conservators with a tool they can use to construct and compare stratigraphies would be extremely helpful as it would facilitate their work in identifying the metal because a comparison would then be possible. There are also many advantages of having a tool that could be used by conservators from all around the world. New corrosion structures could be identified and added to the database. In time, a community of conservators could be created based on this methodology and this tool, which could then evolve over time. The database could then expand with more artifacts and stratigraphies.

Design of a data structure for stratigraphies

In order to support the non-invasive diagnostic of ancient metals we needed to design a suitable data structure that could be used to compare and analyze corrosion products, to support the conservators in identifying the composition of the metal. We started by analyzing the conceptual models used to represent the layers of corrosion products in order to propose a methodology for designing the data structure. We encountered several issues at this first stage: (1) there is no single source of information regarding the various characteristics of the stratigraphies, (2) there are hundreds of characteristics that need to be grouped in order to maintain the overview, (3) there are multiple representations of the same reality due to the fact that the methodology is relatively new and (4) there is currently no exhaustive information on the characteristics of these stratigraphies due to the lack of data.

The first step that we took was to start with a small ontology of the domain, in order to identify the main informational components of an artifact. We identified the 12 concepts listed in Table 1 and displayed as a graph in Figure 4.

Concept	Usage
Artifact	Represents a heritage metal object
Environment	The environment to which the artifact was exposed (fire, water, soil, etc.)
Corrosion type	The corrosion type (Type II, etc.)
Technology	The technology used to craft the artifact
Owner	The current owner of the artifact (museum, etc.)
Object	The main use of the artifact at the time it was crafted (bracelet, etc.)
Origin	The geographic origin of the artifact
Chronology	The estimated period in which the artifact was crafted
Stratigraphy	The representation of all corrosion products of an artifact that could be observed at the same location on the artifact (Bertholon, 2000)
Stratum	One corrosion layer of the stratigraphy (Bertholon, 2000)
Characteristic	A property of a stratum (Bertholon, 2000)
Interface	The junction of two strata from the same stratum (Bertholon, 2000)

Table 1. Main concepts of the ontology representing an artifact and its corrosion layers

Starting with the ontology, we identified the main characteristics of the strata from the available literature. We grouped these characteristics into families of characteristics, each with a description, a selection attribute indicating if the family can be used more than once in each stratum and a list of dependencies (for example the Corrosion Product Stratum Composition can only be used once on a single stratum and the silver metal can be related to chlorine or sulfur). At the end of this cycle, we identified 523 characteristics and their relations. Each characteristic of the model is a member of a family and has relations to other characteristics.

We then considered many different options regarding the methods and the database schemas to choose in order to store the stratigraphies. There are currently two types of database—relational and graph—that can be used to store such a data structure. The results of recent research tend to observe better performance with graphs stored in graph databases rather than in relational databases, even if both ways are well supported by current database systems (Batra & Tyagi, 2012). Graph databases are also more flexible in the case where new relationships need to be added, which is an important property of our own data structure. This led us to choose the Neo4j database (Neo4j, 2012) as the backend for our data structure. When storing data, every node in the database is either an artifact, a stratigraphy, a stratum, an interface (between two strata), a characteristic or a sub-characteristic. These elements are then connected via relationships. The usual pattern is to start from an artifact, which has one stratigraphy made up of several strata. Each of these strata has different characteristics and sub-characteristics, as shown in Figure 5.

The next step was to generate the data structures out of the conceptual model. As we have seen, in order to add information that is not represented in the conceptual model, we have to "supercharge" the model. This allows us to add the necessary elements to generate pertinent data structures. This step was done in collaboration with experts in conservation-restoration. We built a support system in order to aid the selection and application of specific characteristics to strata that were drawn from the stratigraphies. For each stratum, the expert had a choice of families to choose from, based on their relations with the type of stratum. Each time a characteristic is selected and based on its relations with other characteristics, the choice of families is adapted to reflect the current construct. At the end of this step, we had a graph database containing all the stratigraphies of the artifacts that we used as our test sample.

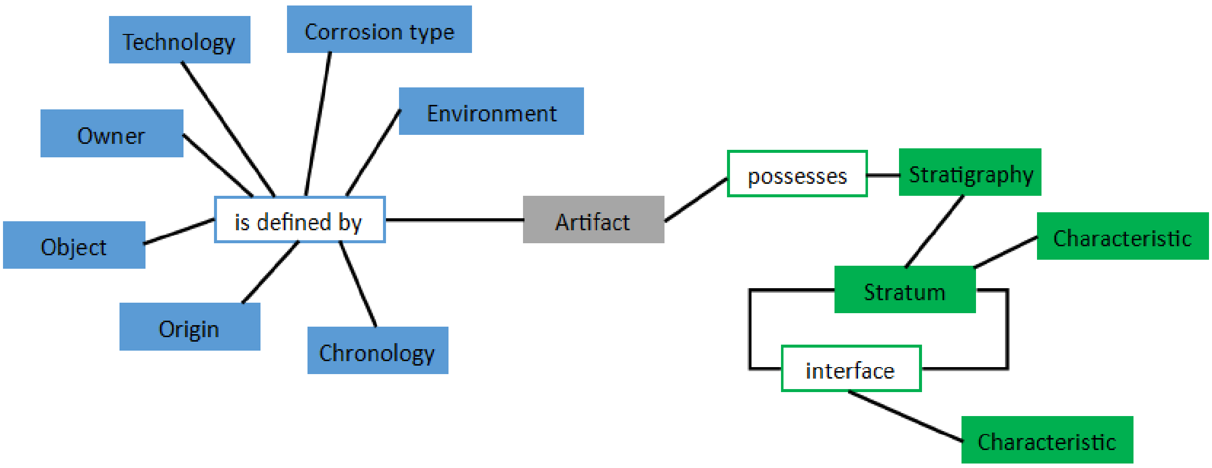

Figure 4. A graph representation of our ontology

We are now able to take a conceptual model of corrosion products (a stratigraphy) and store it in a graph database, supercharged with various families of characteristics, depending on the type of the stratum. The next step was to recall the stratigraphy from the database in order to assess its completeness. There are multiple query languages available to store and retrieve data from a graph database—in our case, the Neo4j database. The preferred approach is to use Cypher ("Cypher Query Language," 2015), a query language specifically developed for Neo4j and offering the developer the same experience as SQL for relational databases (Holzschuher & Peinl, 2013). However, the overall performance of the queries performed with Cypher is currently not as good as using native access to the data provided by the database. In our case, there are no contraindications for using Cypher to query our data structure because the performance reduction compared to native access is negligible for our usage.

Based on the artifacts' characteristics, we created a comparison index, which helped us determine the closest artifacts in terms of structure and relevance for the researchers in conservation-restoration. That index was refined many times after suggested adjustments made by the researchers. With this index, our tool is now able to return the closest stratigraphies to a given stratigraphy, supporting researchers in finding potentially similar artifacts to the one they are studying. However, currently all queries have to be made using Cypher queries, which is not really suitable for our intended audience of conservation professionals.

The final product of the data structure is a digital model presenting all families and their characteristics. It can be used by professional conservators to properly supercharge a stratigraphy with detailed characteristics and by a developer in order to build a rules-based information system that can support the creation of new stratigraphies.

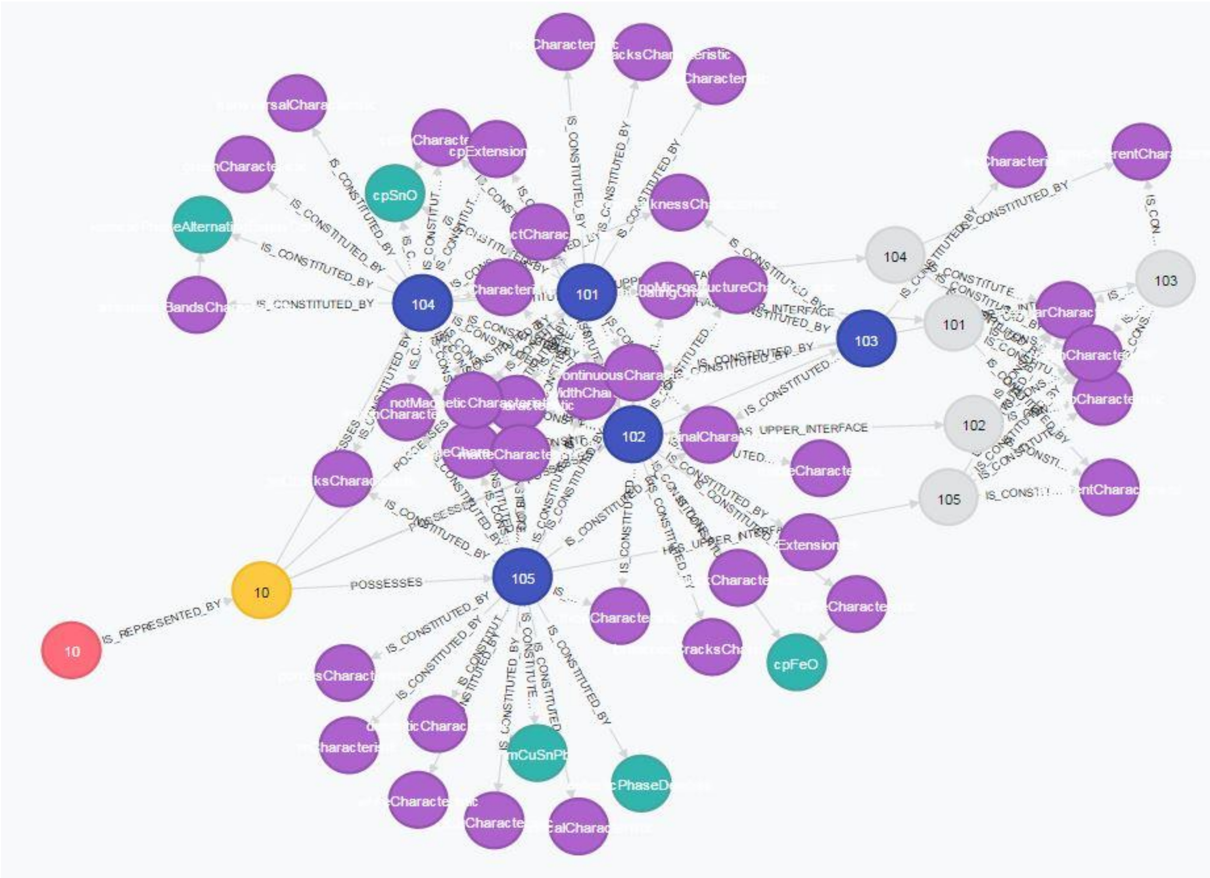

Figure 5. A graph representation of artifact10 in Neo4j

Evaluation of the data structure

In order to evaluate the created data structure, we performed two types of tests. The first set of tests were aimed at checking the validity of the results returned by a comparison query and the second set were performed to assess the completeness of the data structures generated from the conceptual models and backward.

The comparisons of artifacts were performed using Cypher queries on the artifacts that were stored in our database. As seen before, the stratigraphies were created and checked by professionals of the domain. Artifacts data were also completed by information on their exact composition, so that we could assess the results of the comparison queries. We started to run comparison queries on the artifact itself. Hopefully, this returned a 100% matching result. Then we started to search for similar artifacts. Again, performing the query between one artifact and all artifacts stored in the database gave us a 100% matching between the artifact and itself, which was expected.

On the other hand, we found some non-correlated artifacts with exceptionally high matching. After carefully studying the results, we discovered that some typical corrosion products were neglected in favor of highly similar one and resulted in biased results. Common errors were found between artifacts in different environments. Artifacts buried in soil or in water have different corrosion products but can be made of the same metal. However, due to the relatively small sample of artifacts in the current database, we are not able to adapt our comparison queries at the moment. We did not have enough artifacts to identify all discriminant characteristics of our model. Nevertheless, returning more accurate results could easily be done by increasing the weighting we implemented for every characteristic.

In the following table, we compared artifact 16 with the others. As mentioned before, the result of this query returned artifact 16 as the best match. Next, artifact 17 is the second best match. Conservators confirmed that it is indeed the closest artifact to artifact 16 in terms of its composition and structure, which validates our comparison request.

ID	# of strata	Difference # of strata	Total matching	Total relations	Matching100
artifact16	3	0	64	64	**100%**
artifact17	3	0	59	64	**92%**
artifact5	4	1	75	86	**87%**
artifact15	5	2	92	108	**85%**
artifact6	5	2	79	97	**81%**

Table 2. Results of the comparison request with artifact 16

As a result, we are now able to return and sort the artifacts based on the matching with the artifact we want to compare. However, further work has to be undertaken in order to give more accurate results. In the future, we will have more and more stratigraphies in our database, which will help the conservators assess the pertinence of the results. This will also generate more cases in which we will need the conservators' expertise. Therefore, we will have to collaborate closely with them as the database expands.

The second part of the evaluation focused on the information that could get lost in the process from reality to data structure. The first loss is straightforward and always happens when designing a model: as the conceptual model is a simplification of reality, it does not include every aspect of the artifact which the model is based on. Afterwards, in order to have data structures as close to reality as possible, we had to supercharge the conceptual model with characteristics that were not easily understandable or deducible from the drawing. Furthermore, we were able to add annotations to the relationships between an artifact and its characteristics. However, some of these annotations are not taken into account when performing a comparison request. Indeed, a comparison request is only based on the characteristics to return the results and some annotations cannot be translated into characteristics, either because they only apply to a particular artifact or because they are too rare to be implemented in the data structure.

This leads to a second loss of information that occurs in the data structure. When adding an artifact stratigraphy into the database, the conservator has to choose between a limited number of characteristics. Indeed, researchers must provide information about the characteristic family and sub-characteristics for every new addition, whereas they can draw anything on a stratigraphy made by hand. This forces researchers in conservation-restoration to choose between a small range of characteristics and asks them to be very specific and to know exactly what their artifacts are made of. By contrast, they can draw new characteristics on paper, without being absolutely sure about what they are. The main advantage is that this formalism prevents researchers giving different names to the same characteristic. The comparison is therefore easier as the same pattern is followed. However, on our side we have to be as exhaustive as possible concerning the pool of characteristics that the researchers can select from when building their stratigraphies.

Table 3 sums up the completeness of our data structure compared to the conceptual model from the conservators. We can see that we still need to add some characteristics into our data structure to make them available for researchers when building their stratigraphies. Indeed, there are 92 more characteristics that should be in the data structure because they were identified by conservators as such.

		Data structure CR		Total
		Represented	Missing	
Conceptual model CR	Existing	514 (correct)	92 (to add)	**606**
	Non-existing	9 (error)	undefined (future)	**9**
Total		**523**	**92**	

Table 3. Comparison between the conceptual model and the data structure

But there are 9 characteristics that are represented in the data structure but that were not pertinent for the conservators, either because we misinterpreted them or because we made a mistake when adding them to our database. As the database expands, we will certainly find that we have not identified all cases that will occur in the future. Therefore, the main focus will be on the attention paid to the correct identification of the characteristics in order for the data structure to be as accurate and as close to reality as possible. The closer the data structure is to reality, the more accurate the comparison requests will be, as the characteristics will be correctly identified.

Further research

At the moment, as we have to perform Cypher queries to be able to compare stratigraphies, only someone who knows how the Cypher language works and who can interpret the results returned by the comparison requests can make use of our tool. Therefore, we have to develop an intuitive tool that the researchers in conservation-restoration can use, as they are the end users. Of course, the tool must be based on our Neo4j database. This guarantees accurate results when comparing artifacts with one other.

If we want our tool to be used efficiently, we need it to be as close as possible to what the researchers and the students in conservation-restoration are accustomed to. Thus, as they provided us with their conceptual model, we immediately had the idea of using it as the visualization of our data structure. This adds a whole new set of possibilities. Indeed, the user would be able to create a stratigraphy that will be stored into our graph database. Based on its characteristics, the comparison with other stratigraphies will be easy. The user will also be able to visualize the stratigraphy while it is under construction.

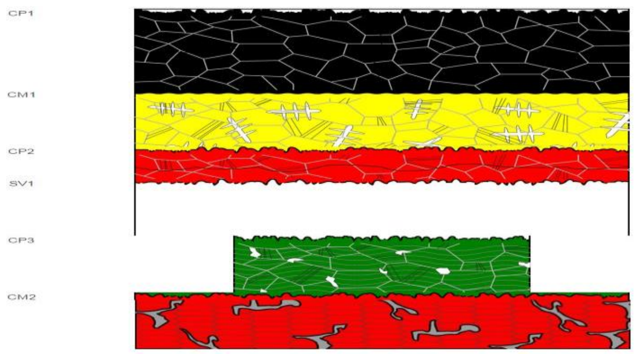

Figure 6. A visualization of stratigraphy #10 using our new tool

This prototype is not yet available for the researchers in conservation-restoration; it still needs improvements to fulfil the needs of the end users. That is why we are closely collaborating with them. Nevertheless, it provides a good idea of what the final product would look like. Visually speaking, we tried to replicate the conceptual models that conservators draw. In addition to that, we translated them so that they could be stored and processed by a computer.

In the coming months, the idea is to make this tool available through a website. That way, people working in the field of conservation-restoration will be able to use it and to supply the database with more data. In the long run, the website should act as a platform which will attract many researchers and students. Then, the latter could consult the database and/or contribute by adding their own stratigraphies.

Conclusion

In our paper, we have talked about the steps needed in order to design data structures starting from a reality that we want to model. In our case, we already had the conceptual model which we had to translate into data structures and then into visualization of these data structures. We realized that the steps we followed were similar to what is done in many other applications.

The most critical point was the translation of the conceptual model into data structures. Indeed, as there was some information that we could not interpret by ourselves and as we needed more information than the ones that are embedded in the conceptual model, we had to collaborate closely with the researchers of the conservation-restoration faculty. This allowed us to supercharge the conceptual model and to come up with relevant data structures that could be further used.

From these data structures, we are now developing a tool that enables the possibility of adding and comparing stratigraphies between them. This would be the first tool available on the market for such an application. Even though the conservators have not started to use it yet, from what we have sounded out, they are determined to do so as soon as the tool is ready.

Eventually, the goal of this project is to facilitate the conservators' work and to provide them with an innovative tool to build their stratigraphies. In addition, this tool is a great way for the methodology developed by Bertholon to be known and used all around the world, as it replicates his stratigraphy model. When the tool is ready, we will have more feedback coming from the field of conservation-restoration. After that, we hope that a community of researchers and students will build up around our tool. Further refinements will then have to be performed, based on their comments.

Acknowledgments

This paper is partly based upon work supported by the RCSO ISnet under Award No. 38996. We would like to thank our colleagues, Christian Degrigny, Régis Bertholon and Romain Jeanneret from the Conservation-Restoration department of the Haute école Arc, who supported our work on this project with their continuous feedbacks on how to improve our decision tool to better fit their needs.

References

Batra, S., & Tyagi, C. (2012). Comparative analysis of relational and graph databases. *IJSCE International Journal of Soft Computing and Engineering*, *2*(2), 509–512.

Bertholon, R. (2000, December 20). *La limite de la surface d'origine des objets métalliques archéologiques. Caractérisation, localisation et approche des mécanismes de conservation.* Université Panthéon-Sorbonne – Paris I.

Cellier, F. E. & Clauss, Christoph, U. A. (2007). Electronic circuit modeling and simulation in Modelica. In *Proceedings of the Sixth Eurosim Congress on Modelling and Simulation* (pp. 1–10).

Cypher Query Language. (2015). Retrieved June 15, 2015, from http://neo4j.com/docs/stable/cypher-query-lang.html

Degrigny, C. & Senn, M. (2012). *MIFAC-Métal: Methodology to Study and Analyse the Microstructures and Corrosion Forms of Ancient and Historic Metals: Application to Metallographic Samples from Swiss Collections*. Neuchâtel.

Holzschuher, F., & Peinl, R. (2013). Performance of graph query languages. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops on - EDBT '13* (p. 195). New York, New York, USA: ACM Press.

Neo4j. (2012). neo4j: World's Leading Graph Database. Retrieved from http://neo4j.org/

Osterwalder, A., Pigneur, Y. & Tucci, C. L. (2005). Clarifying Business Models: Origins, Present, and Future of the Concept. *Communications of the Association for Information Systems*, *16*, 1–25.

Ouyang, C., Dumas, M., Aalst, W. M. P. Van Der, Hofstede, A. H. M. Ter, & Mendling, J. (2009). From business process models to process-oriented software systems. *ACM Transactions on Software Engineering and Methodology*, *19*(1), 1–37.

Peuquet, D. J. & Marble, D. F. (1990). *Introductory Readings In Geographic Information Systems*. Taylor & Francis.

Pidd, M. (2010). *Tools for thinking: modelling in management science* (3rd Edition). Chichester, West Sussex, England: John Wiley & Sons, Inc.

Robinson, S. (2010). Conceptual Modelling: Who Needs It? *SCS M&S Magazine*, 1–7.

Karsten Schmidt, C. S. (2004). A Petri net Semantic for BPEL4WS - Validation and Application. In *Workshop on Algorithms and Tools for Petri Nets* (p. 6). Paderborn: Universität Paderborn.

Wand, Y., Monarchi, D., Parsons, J., & Woo, C. (1995). Theoretical foundations for conceptual modelling in information systems development. *Decision Support Systems*, *15*(4), 285–304.

Wand, Y. & Weber, R. (2002). Research Commentary: Information Systems and Conceptual Modeling? A Research Agenda. *Information Systems Research*, *13*(4), 363–376.

White, S. A. (2004). Introduction to BPMN. *IBM Cooperation*, *2*(0), 1–11.

White, S. A. (2005). Using BPMN to Model a BPEL Process. *BPTrends*, *3*(3), 1–18.