

Chapter 18

Brief History of the Database *Children's Drawings of Gods* (2015–2019)



Olga Serbaeva 

Abstract In 2009, a database (DB) in FileMakerPro was set up for supporting the project *Dessins de dieux* (DDD), also called the *Children's Drawings of Gods* project. In 2015, it was migrated into Structured Query Language (SQL). This presentation shall summarize the major changes the database underwent from 2015 to the beginning of 2019. First, I will present the structural particularities of the MySQL database. Then I will provide a brief description of the technical and architectural solutions that had to be implemented in the new, Resource Description Framework (RDF) database, in order to accommodate the needs of the *Children's Drawings of Gods* research project.

Keywords Databases · MySQL · RDF · Data modeling

In 2009, a database (DB) in FileMakerPro was set up for supporting the project *Dessins de dieux* (DDD), the *Children's Drawings of Gods* project. In 2015, it was migrated into Structured Query Language (SQL). SQL is a programming language used for managing data held in a relational database, or for processing the relational data (Dyer, 2015; Stephens & Russell, 2004; Gilmore, 2010; Cabral & Murthy, 2009; Kruckenberg & Pipes, 2005; West, 2013; Schneller & Schwedt, 2010; DuBois, 2014; Delisle, 2006).¹ It is the best choice for working with structured data, where there are direct or indirect relations between different entities (variables),

¹ SQL is one of the most widely used languages for managing databases. Additional information can be found at: <https://www.mysql.com/>

O. Serbaeva (✉)
University of Lausanne, Lausanne, Switzerland
e-mail: olga.serbaeva@unil.ch

which are placed in different columns of the database. One could easily imagine the data model as a set of columns in Excel that are different but related to each other. The alternatives to SQL are Not Only SQL (NoSQL), a relatively new language allowing to manage less well-structured data (Tiwari, 2012; McCreary & Kelly, 2014), and the Resource Description Framework (RDF), which proposes a significantly different data model. This presentation shall summarize the major changes the *Children's Drawings of Gods* database underwent from 2015 to the beginning of 2019. First, I will present the structural particularities of the MySQL database. Then I will provide a brief description of the technical and architectural solutions that had to be implemented in the new, RDF database, in order to accommodate the needs of the research project.

Before going into technical details, one could ask why a project would need an open access online database at all. Such a database is necessary for the *Children's Drawings of Gods* project² for five reasons.

1. It serves as a stable repository, providing all researchers inside and outside of University of Lausanne (UNIL) access to the exactly same version of the data, which would be impossible to organize via, for example, emails or doc sharing.
2. It makes it possible for the researchers to work on the same image simultaneously, to constitute and share groups of images.
3. It preserves the data, regardless of the composition of the team, and it is thus the first step towards the sustainability of the research data, which is now becoming a norm for research. For instance, the Swiss National Foundation strongly encourages the Data Management Plan that is supposed to function even after the end of the project.
4. In addition, the DB contributes to the Open Science movement; i.e., it allows researchers from all over the world to access and use the collected materials.
5. Finally, the DB is a visual face of the project, serving as an image gallery.

State of the Database in 2015

In 2015, the SQL database, hosting about 2500 images, was a simple and flat structure with only 16 active fields (and 10 linkage-, or supplementary fields, either filled in automatically or not actively used). It did not offer the possibility of storing the data in original languages or non-Latin scripts. The active fields (i.e., those that were constantly used for queries by the researchers of the *Children's Drawings of Gods* project) included the following:

Collection The field entitled collection indicates a historical and/or geographical arrangement of images that also served to distinguish the groups of images based

²The international project, *Drawings of Gods: A Multicultural and Interdisciplinary Approach to Children's Representations of Supernatural Agents*, is also known in French as *Dessins de dieux* (DDD), and referred to in this volume simply as *Children's Drawings of Gods*.

upon the particularities of the tasks given to children. The biggest collection in the database as of 2015 was the UNIL collection. It included the data from Japan, Russia, Switzerland, and Romania.

Cote The cote field is a code of the image, summarizing the information about the image that is most important for the researchers. For example, `ir14_te_m_pmm_08_00_sah` (<http://ark.dasch.swiss/ark:/72163/1/0105/7cYtb6dsSPCYOzsSX536DAE.20181215T035543825Z>), encodes that an image was collected in Iran in 2014 (ir14), in the region of Teheran (te). The representation was drawn by a boy (m), exactly 8 years old (08_00), whose personal name is encoded as “sah.” This image was collected in a public school setting (p) with an additional code (mm), resulting in the 3-letter segment (pmm).³

These codes were very useful for searching and verifying the database info. When the data were migrated to the new database (see below), these codes were retained as the main identifiers for the images. It would have been helpful to exchange the places between gender and school code. This would allow researchers to be able to sort the data by code based on the three parts of geographical code (i.e., country, region, school), and it would structure the child's information sequentially (i.e., gender, age, name). Following this pattern, the code from above ideally should have been `ir14_te_pmm_m_08_00_sah`. The new fields in the new DB allowed a resolution of this problem without modifying the core code-structure.

Longitudinal Code(s) Longitudinal code(s) were those same codes of images linked to one another in a special field for images drawn by the same child in different years. As the names and dates of births were not a part of the data stored online, researches worked outside of the DB to identify the longitudinal images.

Page Orientation The field for page orientation was not filled in by the researches, but was identified automatically by the webpage script by comparing horizontal and vertical size of the image in pixels upon upload. As in the majority of cases A4 sized paper was used for drawings, two options (portrait and landscape), were sufficient to describe the orientation.

Country and Region Country and region were also read by the script from the image code upon upload, and translated into the text format, e.g., “ch” would be interpreted as Suisse/Switzerland, while “te” would stand for Teheran, and “vd” for Canton de Vaud. The textual fields, thus generated, were used to provide a brief description of the image on the webpage, together with the child's gender and age, as well as the year the image was collected (drawing the respective fields of gender, age in years, age in months, and year of collection in the SQL DB).

³At the beginning, the code for school had not three, but only two letters (i.e., `ru08_bo_m_pb_15_01_jam` (<http://ark.dasch.swiss/ark:/72163/1/0105/jzIr8SZQRMq=fKE6fdUmVA0.20180702T190158578Z>)). See the explanation on the institution type field, below.

Institution Type The field for institution type had two options: “r” for religious and “p” for public (secular or lay) institutions (schools, in the majority of cases). This was read automatically by the script from the 3-letter part of the image codes that described the schools (i.e., pmm). Originally, the idea behind the other two letters (mm in the above example) was to assign precise religious identity to the institution (e.g., a secular institution would be indicated by “px” in majority of cases, while religious institution could be marked as “rp”, e.g., religious-protestant). However, this had not been applied to all data (as of 2015, the codes were mostly px and rx), and majority of institutions were lay and public. Therefore, beginning in 2015 the *Children’s Drawings of Gods* team started to use those two letters to mark the exact school. When the database was relatively small, it was unproblematic to have only 2- and not 3-letter code for school (i.e., p/r plus one letter). However, the limits of this approach soon became all too clear. A maximum of only 52 schools could be coded: 26 religious and 26 public. When, in 2014, an Iranian collection of 3000 images from some 60 public schools appeared, immediately the need for restructuring of this 2-letter school code became apparent; and the 3-letter code replaced it as the standard. The change of code necessitated the modification of the research materials, but nothing for tracking such changes originally existed in the old DB. It is, thus, understandable that the scholars responsible for some countries opposed the introduction of the new codes for the schools because it would require massive changes not only in the DB, but also in the personal files of the researchers, and would need additional work from IT personnel. Therefore, the images from Russia, USA, and Romania still have some 2-letter school codes. The additional fields in the new DB, including a field for old or alternative codes, have made this problem irrelevant.⁴

Task Fields The task fields store the instructions given to the participants (translated into English and French). They brought this information directly to the webpage, based on the language chosen by the user (English/French). This was accomplished through inputting text in the old DB, but it was more logical to make it a choice from a limited list in order to reduce the number of typos. This has been rectified in the new DB. As the tasks are rather long, and the subtle changes in three to four lines of text are not directly evident to human eyes, the codes for the different types of tasks were also introduced in the new DB.

⁴The problem with the SQL architecture in the old DB was the fact that the 2- and 3-letter-codes for geographical location and the types of the schools, all had to be unique, i.e., it was impossible to have code “sp” for both Saint Petersburg and Sao Paolo, as these geographical objects would be on the same level, and the script would only read them as Saint Petersburg (the upper line in the code), regardless of the fact that the two objects belong to two different countries. As a result, researchers had to compile a thesaurus of non-repeating codes. They then had to consult their thesaurus each time that a new geographical location or school was introduced, because the proper hierarchical links between geographical entities of level 1 (country) and level 2 (region or city) were missing in the SQL architecture. With the construction of the new DB, a rigid, three-tiered hierarchical geographical location was scripted from the very beginning; it removed the necessity of the thesaurus.

Restatement, Description, Commentary Three textual fields (restatement of the task by the child, description of the drawing by the child, and commentary) were filled in by hand by the researchers. The UTF8⁵ that would allow researchers to type any language in its original script and have a full set of diacritical and punctuation marks, was not implemented in the original SQL DB. As a result, the diacritical marks (such as accents and punctuation), even for the French, were regularly misread on import and export. These signs had to be updated and marked with escape signs in order to secure the unmodified export and display. The implementation of the UTF-based original scripts for Russian, Farsi, Japanese, and Nepalese was completed as the first priority in the construction of the new DB.

Keyword and Questionnaire Among the remaining fields in text format that were supposed to be filled in by the researchers, keywords and questionnaire were both empty. The system of keywords that would have been directly linked to a part of image had been developed on an external platform (Gauntlet), and its tree-like structure was impossible to fit into a single plain text field.⁶

The questionnaire field was supposed to host the information from the transcribed questionnaires, but they were on many pages, of various types, and with many questions. Each type of questionnaire would demand a database for itself. Thus, researchers used the questionnaire field only to fill in the religious identity of the child. This had to be completely revised for the new database (see below).

The import or modification of data triggered the generation of a set of technical fields. These included ID (the order number of the imported image), and page, which was crucial for the internal linkage. This second field was an addition to the code using characters, that would mark the uploaded files as: “-r,” the recto of the image; “-v,” the verso of the image, which included the participant’s description of the image and the task restatement; or “-q,” the separate questionnaire. The questionnaires would often contain multiple pages, but they could only be uploaded one page at a time, in JPG format. So for those questionnaires, the additional codes “-s,” “-t,” and “-u” were used for pages 2, 3, and 4, respectively. Finally, the database would automatically store both the name of the person who uploaded or modified an image or data, and the date of those modifications in corresponding fields (uploaded by user, upload date, edited by user, and edition date).

In order to represent a visual structure of the SQL DB, one should imagine the fields linked to the recto of the image. Thus, any modification of the recto or its code would require reloading of the image and all of its parts (verso, questionnaire),

⁵“UTF-8 is a variable width character encoding capable of encoding all 1,112,064 valid code points in Unicode using one to four 8-bit bytes.” <https://en.wikipedia.org/wiki/UTF-8>

⁶In simple words, it was technically impossible to fit the hierarchically structured keywords within a plain text without inventing a system of tags that would be very difficult for the users to learn and to apply. One would need a separate object in RDF, or a separate table in SQL, for each of the major branches of the keywords tree, and that is precisely what has been done in the new DB.

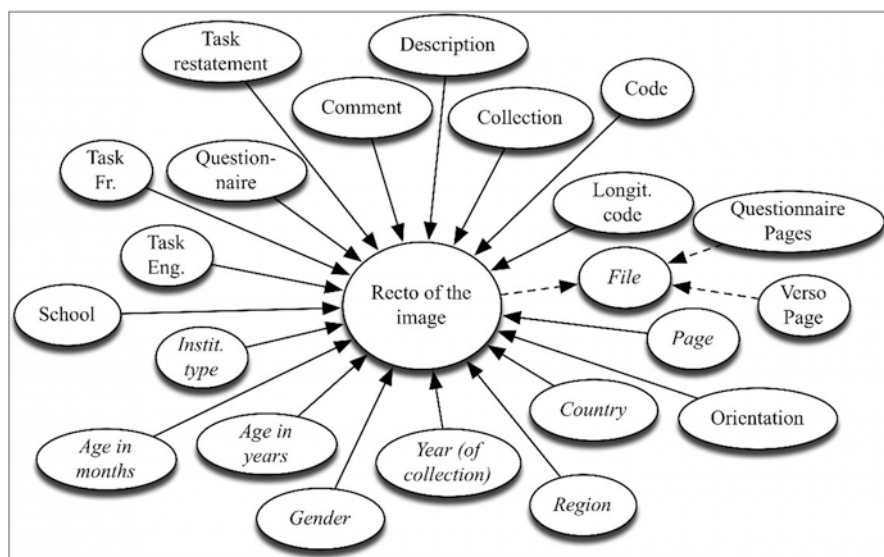


Fig. 18.1 Graphic representation of the SQL database

precisely because of the many script-generated data fields described above. Besides generating data, the script would resize the image to fit various icons in the webpage.

Figure 18.1 provides a graphic representation of the SQL DB, excluding the technical fields.

The fields generated by the script on upload from the code of the recto are in italic. Each oval corresponds to a column in the flat relational database, while each individual drawing generates one line of data in these columns.

The dump of the whole database was approximately 3 MB in 2015. The database allowed the users, via a web-interface, to access the images and the minimal meta-data of the following countries: USA, Japan, Switzerland, Russia and Romania, which constituted the original core of the *Children's Drawings of Gods* research project.⁷

Problems of the Architecture of the SQL Database and Their Influence upon the Research Project

The simple and minimalistic database had the following shortcomings: the data, instead of being entered once per object, had to be repeated for each line of the flat, Excel-like structure. Simply speaking, the design did not fit the basic rules outlined

⁷<https://ddd.unil.ch/>

by Codd (1970, see also Hellerstein et al., 2007).⁸ The existing fields, such as keywords and questionnaire, were not adapted to store the wealth of materials they were intended to hold, and thus remained empty. Simply speaking, each of the existing fields should have been an independent database in itself, with proper links. These shortcomings led to the fact that a lot of the research information could not be stored within the database. Researchers kept this data in external files. This state of affairs not only impoverished the database, but also made the actual data exchange between various parallel research projects more complicated. It became clear that we needed a new database that could integrate, store, and make accessible the majority of the externally stored data.

Meanwhile, the database grew considerably, and in 2017, it included more than 6500 images from the five countries listed above, plus three new ones (Iran, Nepal, and the Netherlands). At the same time, researchers from Greece, China, Argentina, and Brazil were expressing their interest in collaboration.

The architectural problems were dealt with at the same time that the major data standardization process occurred, mostly during 2015–2017. The standardization included the corrections of age in large parts of the collection, and the unification and disambiguation of the codes for schools, regions (for which a standardized, official administrative units-based structure was introduced), and abbreviations of the children's names.

Two Options of Resolving the Structural Problems of the SQL DB

As for the structural problems, there were two possible solutions.

1. The first option involved a series of cosmetic changes to the existing database and the webpage, with the additions of some new fields, but without major restructuring.
2. The second option entailed a total restructuring of the DB.

At the time of this decision, around 2015, various national-level projects aiming at the research data-sustainability had become known. Among them, a potential collaboration with the Data and Service Center for Humanities (DaSCH),⁹ with its Knowledge Organization, Representation, and Annotation (Knora), which is a server application for storing, sharing, and working with humanities data) was envisaged.¹⁰ The DaSCH appeared to be the most promising for the *Children's*

⁸ Summarized here: https://www.ntu.edu.sg/home/ehchua/programming/sql/relational_database_design.html

⁹ Since 2017, DaSCH has become a part of the Swiss Academy of Humanities and Social Sciences. See <http://www.salsah.org/>

¹⁰ <http://dhlab.unibas.ch/knora/> and <https://docs.knora.org/paradox/01-introduction/what-is-knora.html>

Drawings of Gods project in particular, because it was based in Switzerland, run from the University of Basel, and there was a local team in UNIL that would be responsible for the integration.

After a preliminary discussion with the UNIL team of Knora, the *Children's Drawings of Gods* team opted for this second option (i.e., the total data restructuring), in a view that it would lead towards data sustainability and open access. However, the final decision was made and the practical work started when Knora accepted the integration of the restructured DB as a pilot project. This meant that no additional funds were needed for this massive change. The research team agreed to comply with Knora standards concerning the image quality and data structuring, while the Knora team not only committed to the migration the data to the new milieu, but also consented to adapt the generic System for Annotation and Linkage of Sources in Arts and Humanities (Salsah) interface to the precise needs of the *Children's Drawings of Gods* project.¹¹

Opting for the total DB restructuring resulted in the following long-term consequences. First, the database structure had to be rethought from ground zero, both to resolve the problems listed above and to shift from SQL, based on relational data, to semantic query language for databases (SPARQL), which is based on RDF, a model or method for handling data, based on subject-predicate-object relationships, (i.e., triples).¹² RDF proposes a data model which is considerably different from the classical relational database, this difference for humanities, consists in the fact that it allows the integration and analysis of a less-well-structured set of data (as compared to the relational databases, which are extremely rigid, and have limited usability while working with imperfect, incomplete, and/or unstructured datasets). The best overview of the technical side of RDF so far is to be found in Curé and Blin (2015), the pattern search methods applicable to RDF are given in Gerber et al. (2013).

In practical terms, the *Children's Drawings of Gods* project had to shift from an Excel (table-type) data organization to a dynamic industrial standard data model, which had the option of being modified in real time without stopping the database and reimporting all for every object change that one might need to implement. A series of group meetings ensued to discuss the structure. These meetings started in autumn 2016, and the script, with all objects and fields, was ready for migration from SQL to RDF in December 2017. The flat structure of the old SQL database, with, properly speaking, one object with 16 fields, has unfolded into an interconnected net of 18 different objects, with more than 300 fields called *properties*.

The structure of the new database was no longer an internal affair of the members of the research project. From 2016 onwards the project was closely discussed with the Knora/Salsah collaborators, who were also the mediators between the *Children's Drawings of Gods* researchers and the Basel-based main seat of Knora, which had

¹¹ Salsah is a generic user interface (front-end) for Knora (back-end), a software framework for storing, sharing, and working with primary sources and data in the humanities. See <https://dhlabs-basel.github.io/Salsah/>

¹² The main differences between SQL and the SPARQL, used for RDF database queries, are outlined at: <https://www.topquadrant.com/2014/05/05/comparing-sparql-with-sql/>

the final say on the major technical questions. The *Children's Drawings of Gods* team's requests, in that context, could have repercussions both on the generic interface and on the needs of other research projects that were to be integrated into Knora. As a result, some of the team's requests were granted and others, not.

This three-tiered structure (*Children's Drawings of Gods* project—UNIL-based Knora/Salsah group—Basel) has required the establishment of a whole series of communications and organizational tools (meetings, written protocols, Github, sharing data options), and this resulted in significant delays (approximately 1 year) in finalizing the new database and achieving the actual data migration.

The source code for the new database, in its two major parts, ontology and lists, consists of 5511 and 7998 lines of code, respectively.¹³ The migration dump, which included 48 different fields for more than 6500 images and integrated all the fields from the old SQL database plus many of the external Excel files produced by the participating researchers, contained close to 20,000 lines, and occupied approximately 16.5 MB.

The migration started in December 2017. First, the images and the data were imported onto the test platform. After careful testing, they were pushed to production, country by country. Since early February 2019, the new database, accessible at <https://salsah.unil.ch/>, has been fully operational.

The State of the Database in February 2019: The New Database Structure

The new database is about 20 times as big as the old one, with regard to the number of fields.¹⁴ To put it simply, each of the old actively used fields (see above) has become an independent object, a mini-database in itself. In addition, new objects, non-existent in the old database have been created, as well as a complicated network of links between and among them. The graphic representation below represents only the most important links between the 18 objects (see Fig. 18.2).¹⁵

I will now briefly outline the content of the new and upgraded objects of the RDF-based DB.

Instruction This object is based on the Task Eng. and Task Fr. of the SQL DB. The tasks given to children are now accessible in three languages, the original one (i.e., that in which they were spelt out to the children), English, and French. A typology of the tasks, depending on their closeness to the *Children's Drawings of Gods* and

¹³ It is available on GIT: <https://github.com/LaDHUL/drawings-gods>

¹⁴ Available at: <https://salsah.unil.ch/>, select “Dessins des dieux” [drawing of gods] in the vocabulary.

¹⁵ This is a simplified representation of the data model was done with OmniGraffle. The data model itself was created in June 2017, in collaboration with UNIL team of Knora/Salsah, Marion Rivoal and Loïc Jaouen.

UNIL standards has been developed. This closeness is based primarily on the terms used to say *god*,¹⁶ and on the linguistic implications of the formulation concerning the gender and number of gods, that might affect children's drawings. The type of instruction given to the children now largely defines the object called Research Wave (see below).

Person This is a new object created as a response to the problem of linkage of the longitudinal images, i.e., the drawings done by the same child in different years. If in the SQL DB, there was only an internal link between the codes of the rectos; here, the drawings are linked to a Person Object as a larger level entity. It thus becomes possible, for example, to take only the first image done by every child, which would have required a manual sorting in the SQL DB. As this object allows storing the sensitive information, such as the name, the date of birth, the country of origin, the ethnicity of the child, and the spoken languages, it is accessible only to the administrators, but provides a solid ground for the identification of the longitudinal images from within the database.

Directly linked to person object is a set of additional person-linked files, for example, the tests of attachment, etc.

Drawing This object splits into two sub-objects, identical in structure: Public Sub-Object and Private Sub-Object. The Public Sub-Object allows researchers to locate all drawings for which there is an explicit agreement (on the part of the parents) that the images can be made freely accessible online. The Private Sub-Object contains those images for which the parents of the children (often those having medical problems), have agreed to make the drawings accessible only to the members of the *Children's Drawings of Gods* research team.

In the Salsah interface, the Drawing Object, with its two subsections serves as the main anchor for linking the majority of other objects, and puts together many fields from the SQL DB, either directly, or as links to the other objects. Among the new features, there is now the possibility of introducing the imprecise age of the child in a clear manner, keeping track of the history of the code changes thanks to special field, and assigning the hierarchically structured keywords.

Keywords The keywords can be assigned on three different levels: (a) those describing the main god-figure(s) or main motif, when such can be identified, and these are the most detailed, i.e., they have many sub lists; (b) those describing the surroundings of the figure, i.e., general context; and, finally, (c) those used for cases when it is impossible to decide if a particular part of the drawing belongs to the god-figure or to the context. In addition, there are keywords concerning general composition or the layout of the image and the religious identity of the composition

¹⁶Why the term *god* begins sometimes with an uppercase letter G, sometimes with a lowercase letter g, and why it appears sometimes in the singular and sometimes in the plural, is explained in the introductory chapter of this book (Chap. 1, this volume).

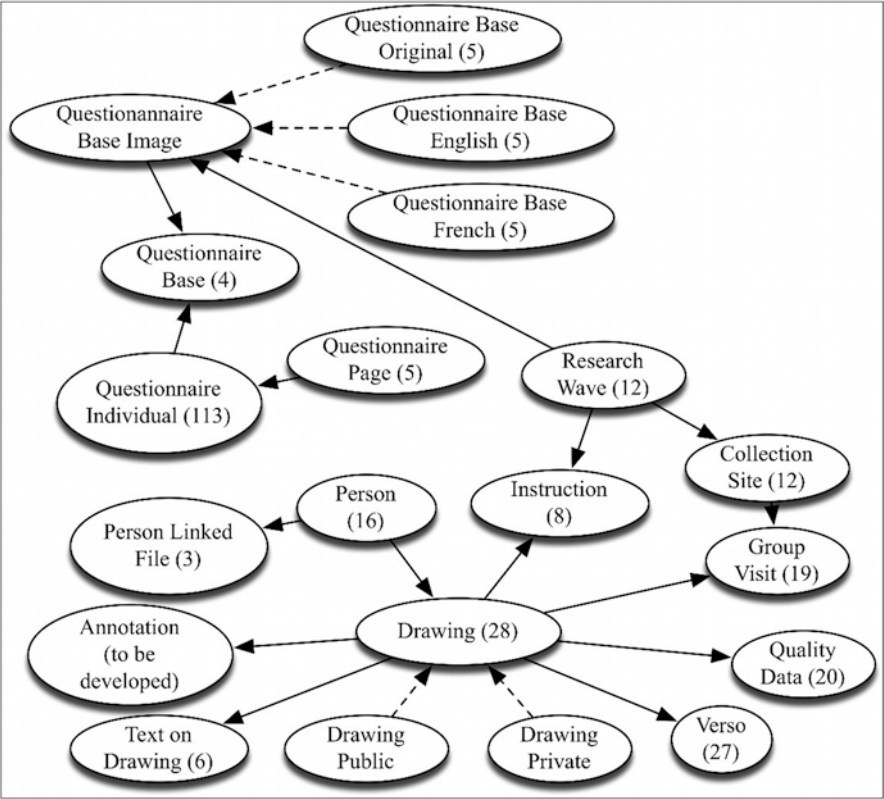


Fig. 18.2 Simplified representation of the links between the objects of the new RDF DB. The number of properties per object appears in brackets. This number excludes the technical properties (date, author of creation, modifications, etc.). Dotted lines indicate a link between an object and its sub-object(s)

In order to provide a better understanding of the hierarchical structure of the Keyword Object and its construction based on the visual analysis of the quasi-totality of the collection (Serbaeva’s working papers, 2016), the following graphic representation has been drawn (see Fig. 18.3).

Within the General Composition area, the researcher can choose from a list to select either a single figure or motif or multiple figures or motifs. In the latter case, an additional list outlines the position of the elements. This is useful to mark the drawings in manga style, for example, or when the drawing is separated into two parts, opposed to each other by the choice of colours and meaning (heaven and hell, etc.)

In the case when the main figure or motif that represents god can be identified, it often falls into one of the following twelve categories, and there is the possibility of adding new categories if such are discovered in the future.

1. Anthropomorphic figure

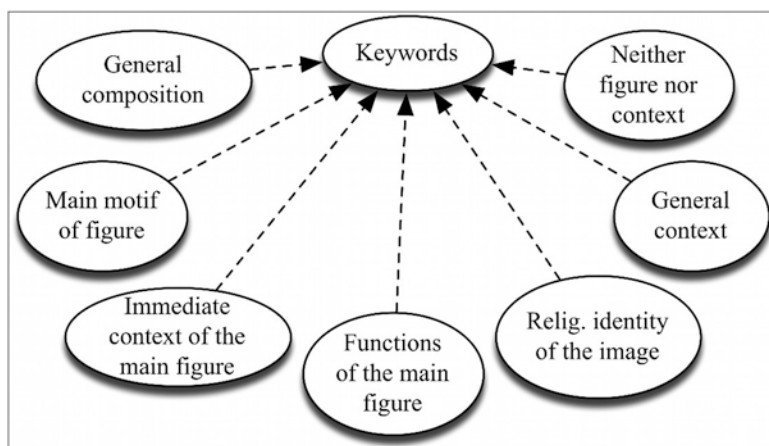


Fig. 18.3 Seven main categories of keywords. Each oval stands for an independent hierarchical list in the new database

2. Passage from one world to another
3. Plant or tree
4. Sun or light
5. Symbol (often having clear religious identity)
6. Text (i.e., citation of sacred writings)
7. Animal, bird, insect
8. Abstract image
9. Clouds, rain, rainbow
10. Cosmos
11. Elements constituting totality (i.e., water-fire-air-earth, etc., in their symbolic representations)
12. Emptiness also including the cases when the child decided to render an empty page

In the absolute majority of cases, i.e., approximately 3500 drawings from 6500, the main figure or motif is an anthropomorphic figure. Each of the above-identified motifs has a number of variants that can be chosen from the list by the researcher. In the graphic representation below (see Fig. 18.4), only the options concerning the anthropomorphic figure will be presented. These options are not exclusive, i.e., more than one can be selected, and, upon selection, they open the additional sub lists.

For example, suppose a researcher would like to use keywords to describe the drawn figure of an angel. The researcher would go to the option Body Particularities and would select Wings and Nimbus from the sub-list there. In case of Naruto, however, the researcher should instead select the option for Recognizable Human Character. Buddha would be classified as Recognizable Religious Character. A

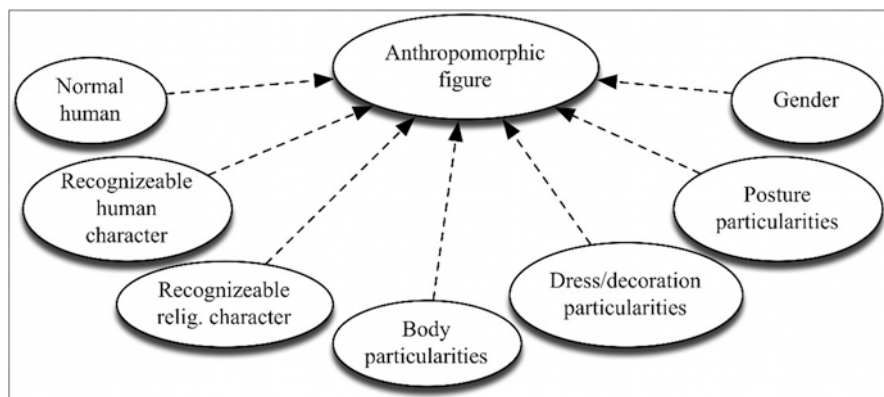


Fig. 18.4 The object, anthropomorphic figure, with seven possible sub-variants. (It is only in very rare cases that one would select the Functions-of-the-Main-Figure and/or the Immediate-Context-of-the-Main-Figure options for non-anthropomorphic images. The religious identity of the image can only be assigned to drawings with clarity when the main motif is a recognizable religious character or when identifiable religious symbols are present, either as the main figure or in the context)

researcher could also note the particularities of posture, or hairstyle in the corresponding categories.¹⁷

The rigid hierarchical structure chosen for the keywords appeared to be the best option, because if the *Children's Drawings of Gods* team had chosen to use a free text option, it would have been difficult to link the synonyms, and one would have to deal with the varied terms and language choices of the participants. For instance, one researcher could describe the angel as “having wings and a nimbus,” while another might write “it flies and has a halo around the head.” Although these descriptions are essentially the same, it would be impossible to apply data mining to such descriptions. Nevertheless, the new RDF DB is structured to allow the introduction of free text comments for every object.

Annotation Directly linked to the drawing is the Annotation object, which is a similar tree of hierarchical keywords, linked to the system of tagging of the actual position of the object in the drawing (measured in pixels). So far, this tagging has been done on a platform (Gauntlet) that is external to Knora/Salsah, and the integration of this data remains an open question (as of Feb. 2019).

Words on the Recto Another object directly related to the drawing object and the set of keywords is the Words-on-the-Recto object. One can tag their presence with the help of the keywords (in main motifs). Nevertheless, in order to transcribe and translate those with multilingual support, an independent object was necessary.

¹⁷The full tree of keywords, including all lists and sub-lists, represents about 17 pages of text, and can be shared on request.

Verso This object represents the textual information from the verso of the drawing, and includes such fields from the SQL DB as task restatement and description by the child. The improvements include the three following aspects: (a) multilingual support, (b) the ability to include the text exactly as it was written by child, as well as a corrected version (if needed), and (c) the possibility to mark who the exact author of a given part of the text was (often in the case of children under 7 years old, the description was written down by an adult).

Questionnaire In the SQL DB this was a reproduction of the pages of the questionnaire in JPG and a single text field was somehow supposed to store the transcribed information. In the RDF DB, the questionnaire field has become a set of objects. This is hardly surprising due to the multiplicity of languages, models, length in pages, and number of questions on the questionnaires. Now every individual questionnaire is linked to its type, and examples of the questionnaires are available for viewing in the original language as it was supplied to the participants, as well as in English and French. The researcher can see the questionnaire in JPG format through the Questionnaire Page object, and enter the data into a single, flat structure based on multiple lists (the same for all questionnaire types) to be found in the Individual Questionnaire object. This simple flat structure, in which one sees all questions that have ever been asked of the participants in the questionnaires of the *Children's Drawings of Gods* project (and which is consequently very long), has been chosen to make the data from various types of questionnaires compatible for the statistical analysis. An alternative option would have been to create a separate object in the RDF DB for each of the questionnaire types, but this would double the number of objects without providing a response to the question of how to make this data compatible across the questionnaires.

Most questions, from the 113 in total, aim at discovering the strength of the religious aspects of the life of a child (how often s/he visits the temples, if s/he prays at home, etc.) They are perfectly compatible and can be used directly for the statistical analysis. However, the questionnaires also contain a lot of hand-written information, which cannot fit in the lists, and for which additional free-text fields have been introduced. Although these additional notes are important to understand the drawings, they are far too random to be the object of data mining.

Research Wave The Research Wave object that has its roots in the SQL DB's collection field is something entirely new. It allows researchers to group the sites where the images were collected in any way they desire. The collections in the SQL DB were often based upon the similarity of tasks and questionnaires, and thus exclusive. This means that the main collection was UNIL; it only included the images corresponding to the *Children's Drawings of Gods* set of standards. Now, however, the researcher can group the images freely, for example, by taking all Catholic schools across the globe, or by selecting a particular language, which could go beyond the limits of a single country. These waves can be assigned to the materials either before or after the data collection.

Collection Site This is an object locating the precise place where the images were collected. In the SQL DB, the corresponding information included country, region, and the coded name of the school. In the RDF DB, the structure is essentially the same, but there are more fields for describing the schools, for example in relation to the gender of the students, or the predominant language. The tripled geo-location structure of this object in the RDF DB allows perfectly tailored access for various kinds of users, i.e., a general user of the database will be able to see only the country and the region, while the team members will have full access to the contact information of a particular site. There are major improvements to the architecture of the data: if in SQL DB the geo-location information was essentially repeated for every image of the same provenance, in the RDF DB the geo-localization is to be entered once only for each collection site.

Group Visit This is a new object that is used to describe the circumstances of each visit to a given collection site. It makes a collection diary accessible, noting the particularities of the age of children in a selected group, or the important circumstances (for example, that “the collection took place after the study of Greek gods,” etc.) Using group visit, researchers can integrate the data of, for example, two different collectors working at the same time on the same site.

Quality Data The last object to be mentioned is also a new one. The necessity of the Quality Data object became apparent during the data cleaning and standardization done in 2015–2017. This object allows the selection of images and metadata that would precisely fit a particular research task. For example, for a color analysis, it makes no sense to include the blank images. Now, thanks to this object, one can select the required data with one click.

The Quality Data Object is also a means of protecting the identity of participating children, who often leave their full names on the rectos of the drawings. These names are cleaned with GIMP¹⁸ and the corresponding field, in quality data, allows the process to be monitored. One can also verify whether or not the child understood the task by comparing the task instructions and the corresponding task restatement. In the RDF DB, the number of fields to fill in for a given image passed 300. The data input is often done by people based on their preferred language, the control of the textual data entry is also done by language, i.e., English, French and, when applicable, the native language of the participant. This simple tool makes it possible to select the images fit for analysis in a situation when the new Database exists, but not all fields have yet been filled in for all images.

To summarize, the RDF DB has better architecture. It is more dynamic, in the sense that one can add the properties and modify the lists without the necessity of

¹⁸ GIMP (GNU Image Manipulation Program) is a free and open-source raster graphics editor used for image retouching and editing, free-form drawing, converting between different image formats, and more specialized tasks. See <https://en.wikipedia.org/wiki/GIMP>

reimporting the data.¹⁹ Now the properties that are common to a group of images need to be keyed only once. The Salsah interface provides excellent search options, both for structured and free text queries, with Boolean search options and wild cards. The access to the data can be defined up to the level of a field within a given object, and the *Children's Drawings of Gods* team can control who sees, reads, and writes in precise sets of objects and fields. The new database handles other languages and scripts (Russian, Japanese, Farsi, etc.), from a technical standpoint, it best suits the present, and the visible future needs of the research project.

Future Development and Open Questions

The new DB, if we revisit the points of the second paragraph of the introductory section, is a stable repository, it definitely contributes to the sustainability of the research data, and it meets Open Science objectives; however, the features of the common work on the same image or a set of images are not well developed in Salsah/Knora's present interface. This interface cannot be adapted to serve as a showcase for the project at this point, as is the case of the webpage linked to the old SQL DB: one can only search and add the information, but one cannot, for instance, acquire a set of random images in order to introduce the DB content to new users. Besides, as the platform stores many databases, it is difficult for new users to find the *Children's Drawings of Gods* project and additional explanations are needed every time.

Other remaining questions to be sorted out with the Knora and Salsah team include export options (not readily available in the present Salsah interface), better integration of the image annotation data (coordinates), and the everyday database enrichment provided by creating an upload script that could prefill the data based on code, (which was a feature of the old SQL DB and linked webpage script).

As the generic interface of Salsah cannot yet be adapted freely to the preferences of the *Children's Drawings of Gods* team, the old SQL database with the linked webpage shall continue to function as the showcase for the project for at least 1 or 2 years after the migration, i.e., approximately until mid-2021.

All that said, the two-party structure that was typical for the research project at the early stages (researchers versus databank), has now become a triangle, as one more player (Salsah/Knora) has been introduced, and this partner is not as reactive and quick as a local IT technician. Besides, the *Children's Drawings of Gods* team cannot implement major changes without having the preliminary agreement from the Salsah/Knora team. If the research project has reached its end, this is not extremely problematic; however, if a new project is put into place based on the same DB materials, a more dynamic and responsive relation between the research team and Salsah/Knora will be necessary.

¹⁹ It is not yet clear whether or not new objects can be added with the same conditions.

References

- Cabral, S., & Murthy, K. (2009). *MySQL administrator's bible*. Wiley.
- Codd, E. F. (1970, June). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377–387.
- Curé, O., & Blin, G. (Eds.). (2015). *RDF database systems: Triples storage and SPARQL query processing*. Elsevier.
- Delisle, M. (2006). *Creating your MySQL database: Practical design tips and techniques*. Packt Publishing.
- DuBois, P. (2014). *MySQL cookbook: Solutions for database developers and administrators*. O'Reilly.
- Dyer, R. J. T. (2015). *Learning MySQL and Maria DB*. O'Reilly Media.
- Gerber, D., Hellmann, S., Böhmann, L., Soru, T., Usbeck, R., & Ngonga Ngomo, A. C. (2013). Real-time RDF extraction from unstructured data streams. In H. Alani et al. (Eds.), *The Semantic Web—ISWC 2013. ISWC 2013. Lecture Notes in Computer Science* (Vol. 8218, pp. 135–150). Springer.
- Gilmore, J. W. (2010). *Beginning PHP and MySQL: From novice to professional* (4th ed.). APress.
- Hellerstein, J. M., Stonebraker, M., & Hamilton, J. (2007). *Architecture of a database system*. NOW.
- Kruckenberger, M., & Pipes, J. (2005). *Pro MySQL*. APress.
- McCreary, D., & Kelly, A. (2014). *Making sense of NoSQL: A guide for managers and the rest of us*. Manning.
- Schneller, D., & Schwedt, U. (2010). *MySQL admin cookbook: 99 great recipes for mastering MySQL configuration and administration*. Packt Publishing.
- Stephens, J., & Russell, C. (2004). *Beginning MySQL database design and optimisation: From novice to professional*. APress.
- Tiwari, S. (2012). *Professional NoSQL*. Wiley India.
- West, A. W. (2013). *PHP and MySQL website databases: A simplified approach*. APress.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

