# HENRI VAN DE WAAL FOUNDATION

Iconographic research, the study of visual culture
and digital Art History

*

# POLICY PLAN 2022-2023

## Objectives

In a world where imagery and the creation of images is becoming more and more important, the importance of the scholarly research of the content of images increases accordingly. That is why the first objective of the *Henri van de Waal Foundation* is to collect visual and textual images and give access to their subject matter.
To this end, it wishes to stimulate iconography as a scientific discipline and to propagate the importance of iconography for the understanding of cultural objects and their historical significance. Due to the large-scale digitization of heritage collections, researchers have been inundated with visual material in recent decades. Classification systems are indispensable for opening up their content. The classification system that is most widely used internationally for this is the Iconclass system, devised by the namesake of the foundation, the Leiden professor of art history Henri van de Waal († 7 May 1972). The second objective of the foundation is therefore to ensure the long-term availability of the online version of Iconclass.
Classification systems such as Iconclass and the Chinese Iconography Thesaurus form an important bridge between the field of the Humanities and that of Information Science - especially in the field of Artificial Intelligence. Therefore, a third objective of the foundation is to bring together researchers from those disciplines and to provide a platform for the exchange of ideas and the initiation of projects.

## Foundation information

The Henri van de Waal Foundation was established on June 14, 2021 and has its registered office in Leiden. The deed of foundation has been filed with notary mr. María de los Angeles Argibary Pérez in Voorschoten.

KvK Chamber of Commerce registration number: 83076573
RSIN: 862716718
SBI-code: 94993
URL: https://henrivandewaalfoundation.org/
E-mail: info@henrivandewaalfoundation.org
IBAN Triodos Bank: NL77 TRIO 0320419916 [ BIC code = TRIONL2U ]
IBAN BUNQ: NL77 BUNQ 2067631667 [ BIC code = BUNQNL2A ]

### Composition of the board (dated 28-12-2021; all members are unpaid)

Chairman: André van de Waal
Secretary & treasurer: drs Hans Brandhorst
Board member: Dr Karin de Wild (Leiden University)

candidate board members:
Dr Pamela Patton, (Index of Medieval Art, Princeton University)
Dr Paul Taylor (Warburg Institute, London)
Dr Hongxing Zhang (Victoria & Albert Museum, London)
Prof Dr Lutz Heusinger (emeritus Marburg University)

# Policy plan 2022-2023

## Summary

In these first years, the Iconclass system will dominate the foundation's activities. In the last months of 2021, a start was made with the new version of the online Iconclass browser, which will replace the current one in early 2022.
The foundation will not only make this new version of Iconclass available, but will also have applications developed that support the use of the system and offer heritage institutions the opportunity to significantly accelerate the scientific disclosure of image material.
To ensure the long-term availability of Illustrated Iconclass and the additional software, the source code and source data of the system are deposited in an open repository (Github).
The availability of code and data will not automatically lead to a stable future for Iconclass. That is why the foundation will also set up a consortium that should financially support the development and technical and editorial maintenance of the Iconclass system. In the course of 2022, international heritage institutions that use Iconclass will be approached to participate in the consortium.
The user registration of the new Iconclass browser will provide an overview of the institutional and individual users of the system.
In addition, information about the system and the foundation will be disseminated via various museum and library organizations and the provision of (online) courses will start.

## Organisation

The appointment of the foreign candidate board members will be completed as soon as their registration with the Chamber of Commerce has been arranged.
In addition, a scientific Advisory Board will be set up. A number of candidates for this have already been approached. It recruits from researchers active in the field of iconography, visual culture and artificial intelligence, supplemented by employees of museums and libraries who have experience in the substantive disclosure of image collections. Recruitment is also made in the publishing world, from members of international organizations in the field of art and culture and from lawyers specializing in intellectual property.

## Scholarly basis

The new edition of Iconclass will be used to underline the scientific importance of systematic iconography.
In recent years it has become clear that the addition of elements of Artificial Intelligence and Pattern Recognition in particular will stimulate the further acceptance of Iconclass. Participation in projects in an advisory role has already become standard practice in 2021. In the coming years, efforts will be made to participate in projects in other ways through training and mediation. This is further explained in the scientific appendix.

## Humanities & Information Science

The need to unlock visual material manually is an inhibiting factor for the substantive study of large amounts of visual data. The foundation now has a large corpus of images made accessible with the help of Iconclass. In the course of 2022, the foundation will compile datasets from this that can be used by researchers from the Humanities and IT disciplines in joint projects. In the past year it has already become clear that there is a demand for this.

## Finance

The Iconclass system and the online Iconclass browser will remain available in Open Access. The institutional users - museums and libraries that make professional use of Iconclass to enrich their collections - will be invited to participate in the consortium. A standard consortium contribution will be requested for this. The importance is obvious: without such a contribution it will eventually become impossible to maintain the Iconclass system and the online system will disappear.
In addition, a program will be set up in consultation with the consortium members and the Scientific Advisory Board for the extension of the Iconclass browser with datasets and supporting applications.
Which extensions and applications are given priority is partly determined by the members of the consortium, who are given a say in the data that will be gathered and the applications that will be developed, in exchange for a premium contribution.
Incoming funds will be used for the further development and application of the Iconclass system and the dissemination of knowledge about it.
In addition, the structural IT and server costs that the websites of the foundation and Iconclass entail will be financed from the consortium contributions and any other income from the foundation.
The application for cultural ANBI (PBO) status has been submitted in February 2022.

## Extensions and applications

For the years 2022-2023, the foundation intends to add the following datasets and applications to the Illustrated Iconclass browser and an iconographic 'toolbox':

### Iconographic bibliography

The existing Iconclass bibliography, more than 50,000 records, can be supplemented and brought up-to-date on the basis of various datasets. A first import, illustrating the principle, is that of data from Kirschbaum's Lexikon der christlichen Ikonographie. But datasets such as the Bildindex or the Rijksmuseum could also provide data for an expansion.

### Expansion of the corpus of illustrations

The Deutsche Digitale Bibliothek - with more than 40 million digital objects, many of which have been described with Iconclass - is just one of the open access data sources that can be used to provide the Iconclass browser with richer illustrations. Public data sources that are not provided with Iconclass, but have been made accessible with controlled vocabularies, are also eligible for import after those vocabularies have been brought into concordance with Iconclass.

### Pattern Recognition Module

The manual assignment of Iconclass concepts to image files is a labour-intensive process. It is relatively slow and expensive. To seriously speed up this process - and perhaps even automate it in part - a module for pattern recognition could be added to the Iconclass browser. That is a long-cherished wish, which could be realized with broader support from a consortium.

### Batch Indexing Module

An additional technique to speed up code assignment could be the simultaneous indexing of groups of images. Batches of images, grouped for visual similarity, could be extracted from large image corpora and simultaneously submitted to an Iconclass assignment module. The matching features or themes could then be indexed in one pass.

### Visualisation of co-occurrences

Datasets with Iconclass codes are potential treasure troves of information about co-occurrences. The images found by a search through one concept are almost always also described with the help of other concepts. Those "co-occurring concepts" can contain a lot of useful information and are a potentially rich source for "knowledge discovery" for researchers. Experience has already been gained in offering co-occurrences, but there is a lot of room for experimentation and improvement, also through visualisation.

### Locally installed browser

Many heritage institutions use Iconclass for the substantive disclosure of their image collections, but do not make optimal use of the search options that the application of Iconclass offers virtually. The Rijksmuseum, the Bildindex and the RKD are just three examples of institutions that have invested a lot in indexing their collections, but do not offer the collected information in an optimized manner - for example via the multilingual keywords - to their end users.
Other institutions mainly use the Iconclass information internally and do not offer it as an option to external users.
A module has been created for this purpose - the Harvester for Iconclass Metadata (HIM) - but its use is currently limited to the Arkyves biotope.
A locally installed variant of this principle could be a welcome addition.

Which of these possible extensions of the Iconclass system will be realized and in which order the applications will be developed, will be determined in consultation with the members of the consortium to be formed. This will also depend on the number of institutions that will participate in the consortium.

## Scientific appendix: image description and pattern recognition

The first goal of the foundation is to facilitate and stimulate the use of the Iconclass system for image research. Visual research methods are widely used in disciplines such as (digital) art history, anthropology and (visual) communication. In these areas, the foundation wants to work closely with universities and heritage institutions.

We are no longer surprised that we can use a picture of a lion instead of the word "lion" to search the Internet for pictures of lions. We are no longer surprised when algorithms such as Google Images present us with series of "visually similar images". We hardly think about how the images presented to us as "similar" were actually selected from the billions of images in Google's databases. And unless we have an IT background, we probably don't understand the technicalities on which the selection is based.


At the same time, most of us do understand that any image, even if it is simply a snapshot of a lion in its natural environment, is part of a narrative. Each image has a context and can be used to tell a story. To tell that story, to document and reflect on the context of an image, we use words. For example, with words we tell that this image is not simply a photo of a lion, but gives us an example of a species whose habitat is threatened by humans.

Images of lions have been used across time and cultures to tell an endless variety of stories and portray many different ideas.


Here is a small selection of the images of lions from the rich corpus available to the foundation. Armed with a little iconographic knowledge you will recognize:

> *the lion as the symbol of the evangelist Mark; the Nemean lion slain by Hercules; a lion as an attribute of Saint Jerome; a lion attacking a monkey depicting a man healing himself from a fever; lions guarding and protecting buildings; the lion killed by Samson; a lion as a symbol of the Power that bows to Justice; a lion portraying a "brave" man; a lion subdued by the love god Amor; a lion as a symbol of Fortitudo and Constantia; a lion as a loving and wise father; Leo, sign of the Zodiac; lions as protagonists in all kinds of fables; the lion as king of the jungle; and a lion that can symbolize both the state and Christ.*

This small anthology of lion iconography shows that recognizing the "pattern" of a lion in the bitmap is only the very first step towards interpreting the image. It is already very complicated for a pattern recognition algorithm to identify lions if we present works of art instead of actual photographs of lions in their natural habitat, to the "eye" of the computer. However, to be able to interpret the images in this overview, we need contextual information that is based on domain knowledge. And contextual information - the word says it all - does not reside in the physical image but in the brain of the human viewer.

Because contextual information changes with the researcher, it is never static and therefore never complete. No matter how much is said about an image, there will always be new ways to look at it, new questions to ask about it, and new information to draw from it.

Documenting the stories that images tell us, and documenting the endless variety of their iconography, requires resources beyond those available to most heritage institutions.
As a consequence, students of the cultural history of images and visual communication are not only confronted with lacunae in their data as many sources are lost to history, but also with disorganized and unsystematic information about the documents that actually *have* survived.

The way images are used to convey a story or a message deserves serious research, now more than ever. Better metadata is an important condition for getting a grip on the content of images, and information technology can play a major role in this because it can help to unlock the content of images faster and better.
The foundation was established to promote visual research and to support the development of software that can take that research to a higher level. It is dedicated to Henri van de Waal because he is not only the creator of Iconclass, but also saw the advantages of a systematic approach to iconography in combination with the possibilities of information technology very early on.

## Case study: the role of an illustrated Iconclass

The Iconclass browser has been in use for more than ten years and is currently being replaced by a new version that is technically up-to-date and also modernized in terms of the language it uses. The indispensable role of ontologies and classification is recognized by both the heritage community and computer scientists working in the field of Artificial Intelligence.
Many museums, libraries and research institutions use Iconclass for content access to their collections and have created numerous datasets with rich subject metadata. When aggregated, they form a large virtual training set for pattern recognition applications. A significant part of these datasets have now actually been merged and they are being used to transform the new Iconclass browser into an "Illustrated Iconclass".
This new edition is of course primarily intended as an authoritative version of Iconclass, but the underlying data can also be used to generate training sets for artificial intelligence applications. Thousands of images, arranged by subject and theme, are a treasure trove from which all kinds of new developments can be drawn. The intention is also to inspire heritage institutions and stimulate software developers to embrace the ideas and build on that.
During the period covered by this policy plan, two of those ideas will already be converted into applications that flank the Iconclass browser via the foundation:

1. The known is used to interpret the unknown. That is, the dataset of classified images will be used as a "knowledge base" for identifying the content of new images presented to the browser application for identification. Pattern recognition is the first step in a supervised learning process. The terms manually assigned to the images are presented as candidate terms to human reviewers.
2. The process of assigning terms to images is accelerated by making it possible to simultaneously classify batches of images with the same or related content.

For a long time, the procedure for cataloguing and retrieving images has seen little variation. A description of the image in free prose or by means of a series of keywords was entered into a database record to unlock its content. Those words were then used to retrieve the image.
The use of a standardized vocabulary such as a classification or a thesaurus does not materially affect that procedure. A standard label such as "*Saint George killing the dragon*" is still found in the vocabulary via keywords and those keywords are also used to retrieve images.
The new Illustrated Iconclass browser will deploy a large corpus of images encoded with Iconclass to offer a second route.

Above is a small selection of the images of Saint George in the corpus administered by the foundation, to illustrate the idea.

All images in that dataset are manually provided with Iconclass labels. To help identify the subject of a print as shown on the right, a reproduction of it should be uploaded.

Its visual characteristics can then be compared with all the images in the corpus. If a match is found via the primary pattern matching, a secondary subject search - using Iconclass - can provide a much richer set of images of saint George for review. The other terms associated with that set - the "*co-occurrences*" - can then also be shown as suggestions for further enrichment or alternative searches.

We are not suggesting that artificial intelligence can automatically identify the subject of an image, its iconographic features or its meaning in this way. On the contrary, the objective is more modest, namely to have an algorithm offer suggestions to a human viewer who can then either accept or reject them. By adding the viewer's judgment to the dataset, the Illustrated Iconclass becomes a "learning" application.

By having the (registered) users assess the suggestions of the algorithm, the quality of the aggregated information increases. For example, an image showing a visually similar pattern, such as the thumbnail on the left, could be annotated by iconographers with the Iconclass term "*Perseus on his winged horse Pegasus, slaying a dragon to liberate Andromeda*", thus correcting "*saint George slaying a dragon*".

In the *Illustrated Iconclass* application, pattern recognition will be combined with textual metadata to speed up the process of filtering visually similar images from the corpus for (further) classification. This enrichment makes use of the fact that by attaching an Iconclass concept to an image, we link that image to a standardized set of (key)words.
For example, the Iconclass label **11H(GEORGE)** links this definition to an image: "*the warrior martyr George (Georgius); possible attributes: banner (red cross on white field), (red) cross, dragon, (white) horse, broken lance , shield (with cross), sword*." In addition, the keywords with which we can find that concept in Iconclass are added to it, so that every image to which 11H(GEORGE) is attached has this "Bag of Words" attached:
*attributes, banner, breaking, broken lance, Christian religion, cross, dragon, George, Georgius, horse, lance, martyr, red cross, religion, shield, soldier, supernatural, sword, warrior, white field*
This bag of words is also supplemented with a comparable set of words from Iconclass's translations.
When a corpus of sufficient size has been processed with Machine Learning algorithms, the application can provide relevant sets of images for identification and enrichment.




For example, these four images could have been judged by a pattern recognition algorithm to be visually similar to Saint George slaying the dragon. Possessing sufficient domain knowledge, however, an indexer would link the images to the Iconclass code **94S32**, which includes a bag of words made up of the theme's definition and a series of keywords:
- *Bellerophon, flying on Pegasus' back, kills the Chimera, a fire-breathing monster, with arrows or a spear*
- *Bellerophon · Chimera Greek legend · Iobates (King) · Lycia · Pegasus · ancient history · arrow · breathing · classical antiquity · deed · fire · hero · heroic legend · history · killing (animal) · king · legend · monster · mythology · spear · task*




Following the same procedure of pattern recognition and object identification, these images of lions as a group would have been filtered from the corpus with this Iconclass label as a suggestion: **25F23(LION)(+1)** to which the following bag of words is associated:
- *beasts of prey, predatory animals: lion (+ animals used symbolically)*
- *animal · earth · lion · mammal · nature · predatory animal · world · symbol*

Two millennia of European art has provided us with many thousands of images of saints, most of which are no longer recognized by modern visitors to museums and churches. It would therefore

be useful if an *Illustrated Iconclass* application could make suggestions for their identification based on image elements such as their attributes. The assignment is simple - at least conceptually: if an algorithm can recognize the pattern of a woman and a tower, then based on previously manually assigned encoding, the application can also suggest that we are dealing with images of Saint Barbara.

The label **11HH(BARBARA)** is associated with a rich bag of words:
*- the virgin martyr Barbara; possible attributes: book, cannon(ball), crown, cross, chalice with wafer, Dioscuros (her father), peacock feather, sword, torches, mason's tools, tower*
*- Barbara (St.) · Christian religion · Dioscuros · Host · cannon · cannonball · chalice · cross · crown · feather · martyr · mason's tools · peacock feather · religion · saint · supernatural · sword · tool · torch · tower · virgin · woman*

Our dataset already contains thousands of images of lions with very different symbolic meanings and thousands of images of saints like this Barbara.
Even if artificial intelligence cannot determine the meaning of a new set of images added to the corpus, we can still expect meaningful suggestions based on the existing data set.
When larger amounts of images can be submitted to pattern recognition algorithms, they can be pre-sorted based on visual features. This alone can seriously speed up the labeling process.

It is also important to realize that Iconclass is not the only way to organize information about images. The *Warburg Institute's Iconographic Database*, the *Princeton Index of Medieval Art*, the *Chinese Iconography Thesaurus*, Getty's *Iconography Authority*, and the Library of Congress *Thesaurus for Graphic Materials* are just a few examples of datasets and vocabulary standards with a similar objective. In addition, countless museums and libraries have developed their own vocabularies, tailored to their collection and their own world view.

At many levels researchers and curators are thinking about how to connect different vocabularies. The foundation's contribution to this process can only be modest, certainly in the period covered by this plan. However, because Iconclass is the most widespread system for iconographic information, its availability in the form of an online browser and in the form of raw data files is an important contribution to the digital Humanities and digital Art History. Depositing the data files and the source code of the browser in the open repository Github is a necessary but not sufficient step.
We can provide an *Illustrated Iconclass*, complete with AI functionality; we can make data files and source code easily accessible to curators, iconographers and software developers and we can make it easy to link the Iconclass concepts to the terms of other vocabulary systems.
None of this will by itself improve the quality of the information about the contents of the collections of libraries and museums, or of the research results in the humanities.

There are good reasons why words such as *encourage*, *promote*, and *facilitate* are used in the objectives of the foundation. Only when what we offer is actually used is it of any use.


Anyone familiar with early modern cultural history will recognize the emblem of an ostrich that spreads its wings but cannot fly.

The message of the English motto "*To have, and not to use the same, is not our glory, but our shame*" may sound too moralistic for modern ears, but like so many emblems, there is much truth in it.

Building a system and making it available, that is within the possibilities of the foundation. Bringing this system to the attention of potential users is what we can also do with this policy plan. Over the next two years, we will continue to build a network of interested colleagues who can help spread this message. However, it is ultimately up to the Humanities and IT communities to decide whether they want to fly or keep walking.