

# The Leonardo Code: Deciphering 50 Years of Artistic/Scientific Collaboration in the Texts and Images of Leonardo Journal, 1968-2018

Clarisse Bardiot ([clarisse\\_bardiot@mac.com](mailto:clarisse_bardiot@mac.com)), Université Polytechnique Hauts-de-France and Peter Broadwell ([broadwell@library.ucla.edu](mailto:broadwell@library.ucla.edu)), UCLA and Mila Oiva ([milaoiv@utu.fi](mailto:milaoiv@utu.fi)), University of Turku and Pablo Suarez ([pablo@im.unam.mx](mailto:pablo@im.unam.mx)), UNAM and Melvin Wevers ([melvinwevers@gmail.com](mailto:melvinwevers@gmail.com)), KNAW Humanities Cluster

[XML](#)

In 1959, C.P. Snow gave an influential lecture on the split between the two academic cultures – the sciences and the humanities (Snow, 1959). Snow deplored the growing divorce between these worlds and their inability to enter into a dialogue. Nonetheless, around the same period, collaborations between art, science and technology began to develop (Debatty and Grover, 2011). Snow's address, which was widely publicized and commented upon (Lee, 2004; Goodyear, 2004), shows that these collaborations cannot be taken for granted. From the 1960s to the present, collaborations between art, science, and technology flourished, leading to the establishment of dedicated cultural institutions and programs, university departments, and academic journals. Among them is *Leonardo* (1968-present), published by MIT Press, which aims to be a forum for interaction, collaboration, and the sharing of ideas between artists, scientists, and engineers across the globe. The spirit of the journal was inspired by the multifaceted activities of Renaissance polymath Leonardo da Vinci, and by the modern utopian vision of bringing together the best minds of humankind. *Leonardo* is the leading international peer-reviewed publication on the relationship between art, science and technology, making it an ideal dataset to analyze the emergence and dynamics of such complex collaborations.

This research studies how interactions between art, science, and technology in *Leonardo* evolved between 1968 and 2018. We use the approximately 3,100 articles and around 7,800 illustrations in the 232 digitized issues of the journal and their metadata as our source data set. Each document is available as a PDF file with full-text searching and an accompanying XML file containing the document's publication metadata.

While text analysis remains dominant, image and other media analysis is an emerging field that may offer unexpected insight, especially where the arts are concerned (Fyfe and Ge, 2018; Wevers and Smits, 2019). This short paper relies on metadata, text (articles, captions) and images (photographs, diagrams, sketches) to offer a fuller examination of the development and interactions networks between communities of researchers and artists in the journal. We applied multiple analytical approaches to the different materials in the dataset. For example, we analyzed illustrations, topical evolution and semantic transformations within the text, and cross-referenced respective results. We also studied the shifting dynamics of the author/institution network interactions collaborations over time.

To study the visualizations, we isolated 7,836 illustrations from all issues of the journal using the

PDF Figures 2.0 open-source package, which is based on PDFBox tool for parsing scholarly publications [1](#). We examined the features of the images in order to determine which modes of representing (and actually doing) art, science or engineering were represented in the images. To this end, we used visual and conceptual clustering based on automatic feature extraction from the images. We ran the PixPlot tool (figure 1) on the images to perform the feature extraction and feature-based similarity clustering; PixPlot uses an Inception3 Convolutional Neural Network, trained on the ImageNet 2012 [2](#) data set, and projects its similarity inference results into a two-dimensional manifold with the UMAP algorithm [3](#) such that similar images appear proximate to one another [4](#). The algorithm allowed us to identify 16 individual clusters that could be considered as lying along a visual genre continuum from “art” to “science,” including archetypes such as technical diagrams, portraits, installations, and sketches. Some clusters are strongly oriented towards “science” or “art,” while others exhibit a mixture of both genre polarities, such as one that weaves kinetic art displays together with more scientific optical imaging devices like radar readouts. Some clustering is representative of periods of time: for example, illustrations from the 1970s exhibit a strong association of technical computer science and engineering diagrams with the notion of “science.” Other features remain constant from the 1950s to the present: for example, portraits of the people involved in the projects, photographs of the installations, standard scientific figures. To understand the longitudinal evolution of our image corpus we will adapt some functionalities of the SIAMESE toolkit [5](#), which visualizes changes in advertisements over time.

As a next step, we will tag the images according to their clusters to determine, as with a latent text topic modeling approach, the proportion of each “class” of images that is present in each paper. This information can subsequently be paired with the information generated through text mining methods. We will use both unsupervised (LDA, NMF) and supervised (Naive Bayes) clustering techniques to determine to what extent a text fell within the arts or science domain [6](#). Moreover, we will extract entities (locations, organizations, people, works of art) and other linguistic features from the text. Pairing the information gathered from the image and text analysis will allow us to answer questions like: in an article describing an art/science collaboration, is it customary to include representative “art” and “science” images, and if so, do these types appear in similar or dissimilar proportions? Also, we will be able to see whether specific actors or networks of actors were responsible for advancing particular visual styles.

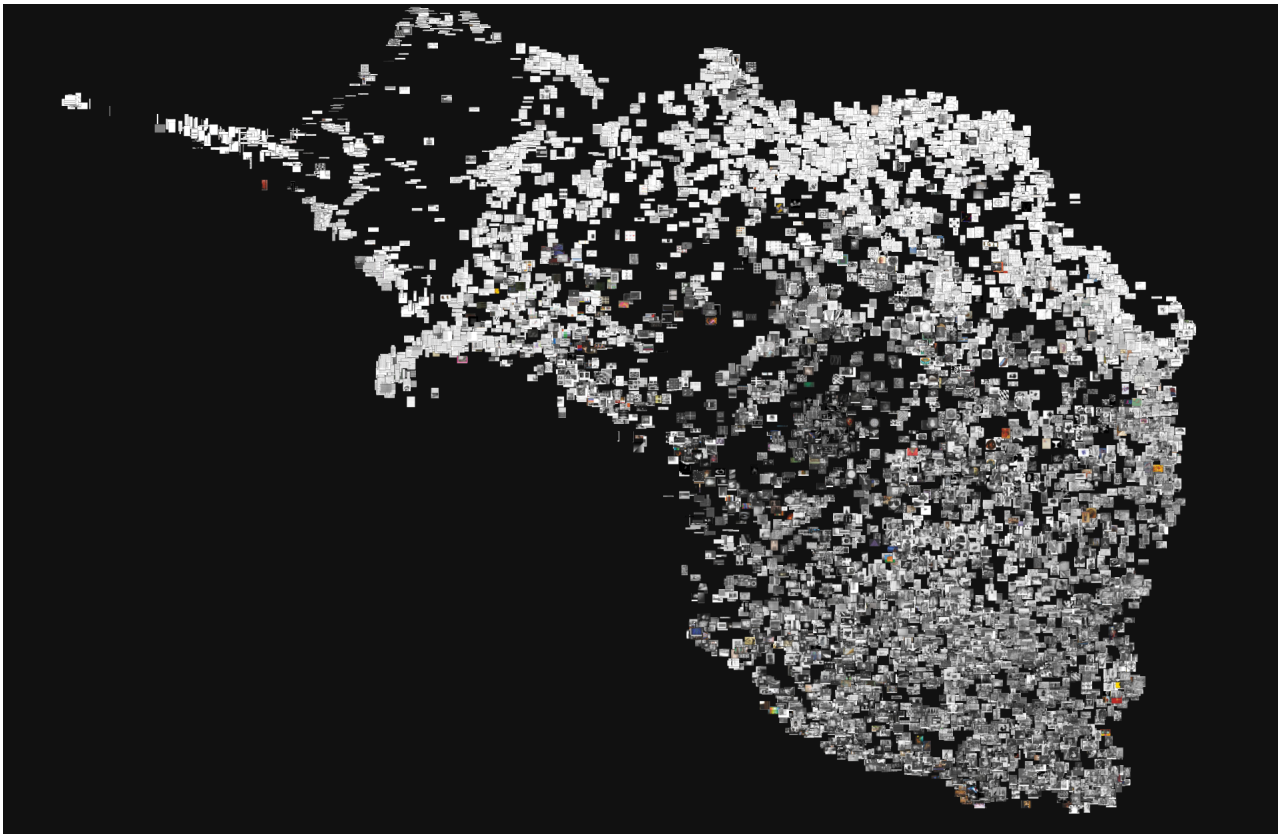


Figure 1: Overview of the *Leonardo* images dataset as visualized via PixPlot.

#### Notes

1.  
<http://pdffigures2.allenai.org/>
  2.  
<http://image-net.org/challenges/LSVRC/2012/>
  3.  
<https://github.com/lmcinnes/umap>
  4.  
<https://github.com/YaleDHLab/pix-plot>
  5.  
<http://lab.kb.nl/tool/siamese#live-demo>
  6.  
SKlearn and Spacy were used for clustering and NLP.
-