

# Introduction, Background, Fundamentals

Suzanne Briet (1894–1989) is recognized for her pioneer role in laying the foundations of modern information science. In her manifesto titled *Qu'est-ce que la documentation?* (“What is documentation?”), she defines a document as *evidence in support of a fact; a physical or symbolic sign, preserved or recorded, for representing, reconstructing or demonstrating a physical or a conceptual phenomenon*. Thus, a document *documents* something, it is a representation serving as evidence for some purpose. We can also say that the information it provides is presented in such a way that the producer of the document becomes able to convey this information to its users/readers, in an intelligible way. This again means that producer and reader use shared representation rules for the information.

Compared to oral communication, the main difference is the preservation of the representation, so that the consumer of the information does not need to communicate directly and immediately with its producer. The most ancient and well-known document type is of course the written document; the introduction of writing was of such importance for mankind that historians usually consider that history starts at that moment, when knowledge of the past can be recovered from written records.

In ►[Chap. 1](#) (A Brief History of Documents and Writing Systems), Henry Baird introduces us to writing systems and briefly reviews the history of written documents. Throughout the centuries, these documents have been physically produced on various materials, using many different writing or printing processes, until this century’s documents which are very often originally available in electronic format. Elisa Barney Smith reviews the evolution of these creation processes and equipment in ►[Chap. 2](#) (Document Creation, Image Acquisition and Document Quality). She also introduces us to the numerous acquisition processes for converting physical documents into images for further processing by appropriate software tools.

When people hear about document image processing, they probably think first and foremost of optical character recognition (OCR). But in the same way as children do not only learn to decipher characters, but progressively learn to understand complete documents, document analysis systems also have gone much further than just OCR. This evolution from character to full and complex documents

is presented by Henry Baird and Karl Tombre in ►[Chap. 3](#) (The Evolution of Document Image Analysis).

A common, fundamental toolbox for all these document analysis systems is that of image processing algorithms applied to document images. Image processing as a whole would deserve a full handbook on its own, but in ►[Chap. 4](#) (Imaging Techniques in Document Analysis Processes), Basilis Gatos presents an overview of the most fundamental image processing methods needed in any document processing and recognition system.