To create others type of data, like text, we use number. Text can be converted to number by using a code. For example: the letter A is 65 in the code, the letter G is 71.  
The American Standard Code for Information Interchange (ASCII) represents all the English letters plus numbers and special characters in seven bits. There is a version of ASCII code that includes other character for German accent for example. But if you speak a language, for example Arabic or Chinese, that does not use English letter, it has been decided to use UNICODE instead of ASCII. UNICODE can represent any writing system and can be “updated” to add other writing system. Older machines still using ASCII, but they’re fading. Most of the stuff represented on a PC is text and numbers, but what about sounds and images?  
Each pixel (picture element) and we don’t know only black and white pictures. The first step to have colors is having shades of gray. It’s possible by using instead of using single bits. High numbers represent light colors (more light), low numbers represent dark color (less light). Brightness is usually represented by 8 bits (0-255). To create color, we need 3 bits (one per primary color: RedGreenBlue). We often use a fourth bit to represent the opacity of a pixel (Alpha). 0 of alpha means a total transparent pixel and a 255 means a total opaque image.   
Once we have images, we can create video by using sequences of images.  
To record sounds, we have to read the pressure variation. To get realistic sounds, we have to check the pressure change very fast, 44100 Hz (41000 times/second). A single check of the pressure is called sample, like a pixel for an image, and the union of samples creates a sound file.

Since an image contains a lot of numbers, a video is usually a file which contains 30/60 images per second. Theatrically, a second of a video occupies an images size x 30. This may fill the memory very fast, so it’s important to store our files in the best way.