TECHNICAL REPORT

ISO/TR 16066

First edition
2003-03-15

Graphic technology — Standard object colour spectra database for colour reproduction evaluation (SOCS)

Technologie graphique — Base de données de spectres de couleurs d'objets normalisée pour l'évaluation de la reproduction des couleurs (SOCS)

Reference number
ISO/TR 16066:2003(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TR 16066 was prepared by Technical Committee ISO/TC 130, *Graphic technology*.

The TC 130 Japanese National Committee prepared this database, and their efforts have made this Technical Report possible. The original form of this report was published as a technical report TR X 0012 by the Japanese Standards Association in Japanese in December 1998.

Introduction

The simplest way to evaluate the colour reproduction of colour image input devices is to input images of objects whose colours are exactly known and then to compare the pixel values to exact ones. For that purpose, input colour target standards have been already established in ISO 12641:1997, *Graphic technology — Prepress digital data exchange — Colour targets for input scanner calibration*. Evaluation becomes more complicated, however, when we consider metamers.

The perceived colours of a pair of objects are referred to as metamers if, under a particular illumination, they have the same tristimulus values even though they are spectrally different. The spectral pairs of such objects may be used to advantage in the assessment of differences among lighting conditions. If, for instance, a pair of spectra yield the same tristimulus values *X, Y, Z* under CIE illuminant D50, the difference between the two perceived colours of the pair of spectra as measured in the field under a different illumination is referred to as a metameric index, which can be used as a measure of the non-conformance of that illumination to D50.

While it may be sometimes preferable, e.g. for standards purposes, to use artificial pairs of perceived colours, i.e. those not coming from natural objects, natural metameric pairs have the advantage that one can assess the effect of non-standard lighting for a particular lighting condition. One may find, for instance, that a particular light source leads to unacceptably large deviations in skin tones, whereas the same source is quite acceptable for furniture colours.

Definitions with respect to metameric indices and the procedures for their evaluation are described in the following CIE publications:

CIE 15.2, *Colorimetry*, 2nd ed. (1986) (Contains Special Metamerism Index: Change in Illuminant)

CIE 51.2, *A method for assessing the quality of daylight simulators for colorimetry* (1999)

CIE 80, *Special metamerism index: Change in observer* (1989)

It can also be useful to consider metamers in the evaluation of such colour image input devices as colour scanners and digital cameras, which, though designed to capture images in a way similar to that of the human visual system (HVS), nonetheless deviate enough from HVS sensitivity so that colour reproduction of sensed colours in display devices or print outputs are significantly different from that desired, even when the illumination conditions of the original human observation of an object have been recreated for the observation of the output image.

To evaluate deviations due to variations in light sources and/or sensor sensitivities under actual conditions, it is useful to know the range of spectral differences in existing objects. Committee members have created an exhaustive collection of colours of existing objects, a database containing more than 50 000 items. This report details the extraction from the data of 365 colour samples and their classification into sets, of which there are two types, “typical sets” and “difference sets”. “Typical sets” refers to sets of typical spectral reflectances and transmittances of objects as classified into a number of different categories. “Difference sets” refers to sets of metamers whose tristimulus values are roughly typical but whose spectral values are significantly non-typical.

The entire original collection of more than 50 000 spectral data items is included, in electronic form, as part of this Technical Report in the data directory SourceData as described in Annex A.

Graphic technology — Standard object colour spectra database for colour reproduction evaluation (SOCS)

1 Scope

This Technical Report provides a database of typical and difference sets of existing object colour spectral data that are suitable for evaluating the colour reproduction of image input devices. It also includes the spectral reflectance and transmittance source data from which these data sets have been derived.

2 Standard object colour spectra

2.1 Object categories and sample selection

2.1.1 Categories and groups

The following categories and subcategories were first established:

(1) Photographic materials

a. Transparencies

b. Reflection prints

(2) Offset prints

(3) Computer colour prints

a. Dye sublimation printer

b. Electrostatic printer

c. Ink-jet printer

(4) Paint (not for art)

(5) Paints (for art)

a. Oil paints

b. Water colours

(6) Textiles

a. Synthetic dyes

b. Plant dyes

(7) Flowers and leaves

(8) Outdoor scenes (Krinov data except for flowers and leaves)

(9) Human skin

a. Bare North Asian skin

b. Foundation-applied North Asian skin

c. Bare South Asian skin

d. Foundation-applied South Asian skin

e. Bare Caucasian skin

f. Bare Negroid skin

Spectral reflectance/transmittance data were then collected for more than 50 000 items falling into these categories/subcategories.

Categories (1) to (5) and subcategory (6)a are for artificial colours, while subcategory (6)b and categories (7) to (9) are for natural colours. Typical sets and difference sets were established from 365 samples in this database. A typical set is a set of representative spectral data of colour objects, while a difference set is a set of metamers whose colour under D65 illuminant is similar to typical set samples but differs significantly from them spectrally. Samples for the typical and difference sets were selected as shown in Table 1. The manner of selecting typical samples depended on whether colours were artificial or natural. In most artificial colour groups, all colours are synthesized by mixing three or four colorants, and distributed almost uniformly in their colour gamut. Colours in the paint (not for art) and paints (for art) categories, however, are synthesized by mixing more than four colorants, and these categories were dealt with in the same manner as with natural colours. Selection strategies are described in following subsections.

Table 1 — Numbers of selected typical/difference set colour samples

Group	Typical sets	Difference sets
Photo (transparency)	15	15
Photo (reflection print)	15	15
Offset prints	15	15
Dye sublimation printer	15	15
Electrostatic printer	15	15
Ink-jet printer	15	15
Textiles (synthetic dyes)	15	15
Flowers/grasses/leaves (includes Krinov's grasses and leaves)	25	25
Paint (not for art)	15	—
Oil paints	15	—
Water colours	15	—
Textiles (plant)	15	—
Non-grass/leaf Krinov	15	—
Bare North Asian skin	5	—
FD-applied North Asian skin	5	—
Bare South Asian skin	5	—
FD-applied South Asian skin	5	—
Bare Caucasian skin	5	—
Bare Negroid skin	5	—
Total	**365**	

NOTE 'Foundation' is a cosmetic used as a base for facial make-up. However, in this Technical Report, 'foundation applied skin' means skin that is not bare, but covered with foundation and/or face powder.

2.1.2 Typical set selection for artificial colour groups

There are seven artificial colour groups in Table 1: photographic transparency, photographic reflection prints, offset prints, dye sublimation printer, electrostatic printer, ink-jet printer, and textiles (synthetic dyes). Colour samples can be obtained for every hue in these groups.

Colorants can be expected to vary within any one group, as, for example, among the many products of different photographic prints material manufacturers. Spectral reflectance measurements were carried out for several representative products among them. Statistical analysis was carried out on the measurement data, and the product whose characteristics most closely approximated the statistical average for the products as a whole was determined to be a typical colour product. A mathematical explanation for this is found in Annex B. From a large number of colour samples for a typical product, fifteen samples were selected whose colours are nearest to pre-determined basic colours, which consist of three achromatic colours and twelve chromatic colours that are homogeneously distributed in twelve hues. Lightness and chromaticity of each chromatic colour was such that the colour gamut of each artificial colour group contained all the chromatic colours. Table 2 shows the achromatic colours (1-3) and the chromatic colours (4-15). The colour for each spectral data is calculated under D65 illuminant.

Table 2 — Basic colours for artificial colour groups

	*L**	*H**	*C**	*a**	*b**
1	20	—	0	0,0	0,0
2	50	—	0	0,0	0,0
3	80	—	0	0,0	0,0
4	40	0	30	30,0	0,0
5	45	30	35	30,3	17,5
6	50	60	37	18,5	32,0
7	60	90	45	0,0	45,0
8	60	120	30	–15,0	26,0
9	45	150	30	–26,0	15,0
10	45	180	23	–23,0	0,0
11	45	210	22	–19,1	–11,0
12	45	240	20	–10,0	–17,3
13	40	270	20	0,0	–20,0
14	35	300	27	13,5	–23,4
15	40	330	30	26,0	–15,0

2.1.3 Typical set selection for non-skin colour, natural colour groups

There are six natural colour groups in Table 1 for non-skin colours: flowers/grasses/leaves (including Krinov's grasses/leaves), paint (not for art), oil paints, watercolours, textiles (plant), and non-grass/leaf Krinov. Samples in natural colour groups are not distributed in whole hues, and typical samples cannot be selected on the basis of their colours. To select typical set samples for natural colour groups, an algorithm based on spectral distribution was developed. When principal component analysis was applied to all samples in a group, it became possible to express the data distribution in a low dimensional subspace. An equi-distanced lattice was set in the subspace, and a representative sample was selected from each lattice point. A mathematical explanation for this is found in Annex C. Twenty-five typical samples were selected from the 'flowers/grasses/leaves' group, while 15 typical samples were selected from each of the other groups.

2.1.4 Typical set selection for skin groups

There are six skin colour groups in Table 1: Bare North Asian skin, FD (foundation)-applied North Asian skin, Bare South Asian skin, FD-applied South Asian skin, Bare Caucasian skin, and Bare Negroid skin. Skin colours are not distributed over a wide range, and only five samples were selected from each group. These colours correspond to average and extreme colours over their distribution in CIELAB space. The mathematical explanation for this is found in Annex D.

2.1.5 Difference set selection

Difference set samples were selected for typical samples in the artificial colour groups and in the flowers/grasses/leaves group. Colour samples in difference sets have colours that, while similar to typical set sample colours under D65 illuminant, are significantly different spectrally. Samples whose colour was least different from a typical set were selected first. The most spectrally different colours among these were then selected to create the difference set. The mathematical explanation for this is found in Annex E.

2.2 Typical set samples and difference set samples

Typical set spectral data and difference set spectral data are stored in electronic form as files in the subdirectories designated ‘typical’ and ‘difference’ described in the directory TRDatabase that is part of this Technical Report. All spectral data are 31 dimensional, from 400 nm to 700 nm at 10 nm interval. The spectral data are presented as percent reflectance factor or percent transmittance factor. File names corresponding to the above groups are summarized in Table 3. The originally collected data can also be accessed from the data directory SourceData which is described in Annex A. Correspondence with original sample identification numbers (IDs) is summarized in Annex F.

The files ‘photo_t’, ‘photo_r’, ‘offset’, ‘print_ds’, ‘print_es’, ‘print_ij’, and ‘textiles_s’ in ‘typical’ directory include typical set samples from artificial colour groups, each sample with a number from 1 to 15 has a colour similar to its corresponding basic colour. The same holds for the difference set colours in the files ‘photo_t-d’, ‘photo_r-d’, ‘offset-d’, ‘print_ds-d’, ‘print_es-d’, ‘print_ij-d’, and ‘textiles_s-d’ in the ‘difference’ directory, since they are samples in artificial colour groups. For the other typical set samples and for the other difference set samples, which belong to natural colour groups, there is no such relation to basic colours. Nevertheless, the order of colour samples in ‘flowers_leaves’ file in the ‘typical’ directory corresponds to the order of colour samples in the ‘flowers_leaves-d’ file in the ‘difference’ directory. Such correspondences can be confirmed by comparing the L^*, a^* and b^* values of the samples in Annex F.

Table 3 — File names that contain typical set samples and difference set samples for each group

Kind of set	Group	Directory\File name
Typical set	Photo (transparency)	typical\photo_t.txt
Typical set	Photo (reflection print)	typical\photo_r.txt
Typical set	Offset prints	typical\offset.txt
Typical set	Dye sublimation printer	typical\print_ds.txt
Typical set	Electrostatic printer	typical\print_es.txt
Typical set	Ink-jet printer	typical\print_ij.txt
Typical set	Textiles (synthetic dyes)	typical\textiles_s.txt
Typical set	Flowers/grasses/leaves (incl. Krinov's grasses and leaves)	typical\flowers_leaves.txt
Typical set	Paint (not for art)	typical\paint.txt
Typical set	Oil paints	typical\oil.txt
Typical set	Water colours	typical\water.txt
Typical set	Textiles (plant)	typical\textiles_p.txt
Typical set	Non-grass/leaf Krinov	typical\n_krinov.txt
Typical set	Bare North Asian skin	typical\n_asian_b.txt
Typical set	FD-applied North Asian skin	typical\n_asian_f.txt
Typical set	Bare South Asian skin	typical\s_asian_b.txt
Typical set	FD-applied South Asian skin	typical\s_asian_f.txt
Typical set	Bare Caucasian skin	typical\caucasian_b.txt
Typical set	Bare Negroid skin	typical\negroid_b.txt
Difference set	Photo (transparency)	difference\photo_t-d.txt
Difference set	Photo (reflection print)	difference\photo_r-d.txt
Difference set	Offset prints	difference\offset-d.txt
Difference set	Dye sublimation printer	difference\print_ds-d.txt
Difference set	Electrostatic printer	difference\print_es-d.txt
Difference set	Ink-jet printer	difference\print_ij-d.txt
Difference set	Textiles (synthetic dyes)	difference\textiles_s-d.txt
Difference set	Flowers/grasses/leaves (incl. Krinov's grasses and leaves)	difference\flowers_leaves-d.txt

3 Use of the colour spectra database

3.1 Use of typical sets

In this section examples of how the database may be used to evaluate the colour reproduction accuracy of a device are given. The usual colour reproduction evaluation scheme is as follows:

Object colour is calculated using colour matching functions $(\bar{x}, \bar{y}, \bar{z})$ where $\bar{x} = \left[\bar{x}(400), \bar{x}(410), ..., \bar{x}(700)\right]^t$, $\bar{y} = \left[\bar{y}(400), \bar{y}(410), ..., \bar{y}(700)\right]^t$, $\bar{z} = \left[\bar{z}(400), \bar{z}(410), ..., \bar{z}(700)\right]^t$ recommended by CIE 15.2. If the i-th object's spectral reflectance is $\beta_i = \left[\beta_i(400), \beta_i(410), ..., \beta_i(700)\right]^t$ and illumination intensity is $S = \left[S(400), S(410), ..., S(700)\right]^t$, the CIE-1931 XYZ values for the i-th object may be expressed as in equation (1):

$$\begin{cases} X_{0i} = k \sum_{\lambda=400}^{700} S(\lambda)\beta_i(\lambda)\bar{x}(\lambda) \\ Y_{0i} = k \sum_{\lambda=400}^{700} S(\lambda)\beta_i(\lambda)\bar{y}(\lambda) \\ Z_{0i} = k \sum_{\lambda=400}^{700} S(\lambda)\beta_i(\lambda)\bar{z}(\lambda) \end{cases} \tag{1}$$

where $k = \dfrac{100}{\sum_{\lambda=400}^{700} S(\lambda)\bar{y}(\lambda)}$

NOTE CIE 15.2 recommends that calculation shall be carried out from 380 nm to 780 nm at an interval of 5 nm. However, in the practical cases for which this data is used (with a data range of 400 nm to 700 nm at an interval of 10 nm) it is recommended that weighting functions such as those defined in ASTM E-308 be used.

Colour image input devices usually have sensors with three different spectral sensitivities. If actual sensor sensitivities are $(\bar{s}_r, \bar{s}_g, \bar{s}_b)$ $\bar{s}_r = \left[\bar{s}_r(400), \bar{s}_r(410), ..., \bar{s}_r(700)\right]^t$, $\bar{s}_g = \left[\bar{s}_g(400), \bar{s}_g(410), ..., \bar{s}_g(700)\right]^t$, $\bar{s}_b = \left[\bar{s}_b(400), \bar{s}_b(410), ..., \bar{s}_b(700)\right]^t$ output signals (R_i, G_i, B_i) for the i-th object may be expressed as in equation (2):

$$\begin{cases} R_i = \sum_{\lambda=400}^{700} S(\lambda)\beta_i(\lambda)\bar{s}_r(\lambda) \\ G_i = \sum_{\lambda=400}^{700} S(\lambda)\beta_i(\lambda)\bar{s}_g(\lambda) \\ B_i = \sum_{\lambda=400}^{700} S(\lambda)\beta_i(\lambda)\bar{s}_b(\lambda) \end{cases} \tag{2}$$

The (R_i, G_i, B_i) is converted to colour values in the CIE-1931 XYZ colour space. Equation (3) is often used for the conversion — though a higher order function may sometimes be desirable.

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \tag{3}$$

Matrix elements $\{a_{ij}\}$ can be determined using the least square method, where the sum of squared colour differences for typical set colours are minimized in some colour space. The sum cannot be zero in practical cases. This residual colour difference is evaluated in a uniform colour space (e.g. CIELAB space). The conversion from the XYZ space to the CIELAB space is also described in CIE 15.2. Letting the converted values from (X_{0i}, Y_{0i}, Z_{0i}) and (X_i, Y_i, Z_i) be $(L^*{}_{0i}, a^*{}_{0i}, b^*{}_{0i})$ and $(L^*{}_i, a^*{}_i, b^*{}_i)$, respectively, the average colour difference for n typical set object colours may be expressed as in equation (4):

$$\Delta\bar{E}^*_{ab} = \frac{\sum_{i=1}^{n} \sqrt{\left(L^*_i - L^*_{0i}\right)^2 + \left(a^*_i - a^*_{0i}\right)^2 + \left(b^*_i - b^*_{0i}\right)^2}}{n} \tag{4}$$

The smaller the $\Delta\overline{E}^{*}_{ab}$, the better the colour reproduction. That is, in the event that sensor spectral sensitivities of colour image input devices can somehow be measured, it is possible to evaluate their colour reproducibility using actual object colours. It is also possible to estimate colour reproducibility using this database when designing the spectral sensitivities of image input devices.

Usually, a single image input apparatus does not need to input all the object colours in the world. Film scanners, for example, need only to scan colour films; it is not necessary for scanners to reproduce the colours of natural objects or scenery, and the colour reproduction of colour scanners only needs to be evaluated for colour films. That is why data here have been classified into object groups. For a scanner, therefore, only the typical set samples identified as "photo" in Table 3 (i.e. photo_t.txt and photo_r.txt) would be used for evaluation.

On the other hand, for evaluating digital cameras, human faces, flowers, leaves and outdoor objects always have high importance, and typical sets for these object groups would be used.

3.2 Use of difference sets

Difference sets contain samples whose appearance is similar to the appearance of typical sets samples under the standard illuminant D65, but whose spectral reflectance/transmittance is significantly different from that of typical set samples. That is, difference sets contain metamers. Let us first assume that β_j represents the spectral reflectance of a given difference set sample, a reflectance that corresponds to β_i, the spectral reflectance of a typical set sample. In this case, two sensor outputs will be calculated using two light sources S_1 and S_2. Four sets of R, G, and B values can then be obtained, as expressed in equations (5) through (8).

$$\begin{cases} R_{i1} = \sum_{\lambda=400}^{700} S_1(\lambda)\beta_i(\lambda)\overline{s}_r(\lambda) \\ G_{i1} = \sum_{\lambda=400}^{700} S_1(\lambda)\beta_i(\lambda)\overline{s}_g(\lambda) \\ B_{i1} = \sum_{\lambda=400}^{700} S_1(\lambda)\beta_i(\lambda)\overline{s}_b(\lambda) \end{cases} \quad (5)$$

$$\begin{cases} R_{i2} = \sum_{\lambda=400}^{700} S_2(\lambda)\beta_i(\lambda)\overline{s}_r(\lambda) \\ G_{i2} = \sum_{\lambda=400}^{700} S_2(\lambda)\beta_i(\lambda)\overline{s}_g(\lambda) \\ B_{i2} = \sum_{\lambda=400}^{700} S_2(\lambda)\beta_i(\lambda)\overline{s}_b(\lambda) \end{cases} \quad (6)$$

$$\begin{cases} R_{j1} = \sum_{\lambda=400}^{700} S_1(\lambda)\beta_j(\lambda)\overline{s}_r(\lambda) \\ G_{j1} = \sum_{\lambda=400}^{700} S_1(\lambda)\beta_j(\lambda)\overline{s}_g(\lambda) \\ B_{j1} = \sum_{\lambda=400}^{700} S_1(\lambda)\beta_j(\lambda)\overline{s}_b(\lambda) \end{cases} \quad (7)$$

$$\begin{cases} R_{j2} = \sum_{\lambda = 400}^{700} S_2(\lambda)\beta_j(\lambda)\bar{s}_r(\lambda) \\ G_{j2} = \sum_{\lambda = 400}^{700} S_2(\lambda)\beta_j(\lambda)\bar{s}_g(\lambda) \\ B_{j2} = \sum_{\lambda = 400}^{700} S_2(\lambda)\beta_j(\lambda)\bar{s}_b(\lambda) \end{cases} \tag{8}$$

To compare colour differences, CIELAB values for these four colour values can be calculated as $\left(L^*_{i1}, a^*_{i1}, b^*_{i1}\right)$, $\left(L^*_{i2}, a^*_{i2}, b^*_{i2}\right)$, $\left(L^*_{j1}, a^*_{j1}, b^*_{j1}\right)$ and $\left(L^*_{j2}, a^*_{j2}, b^*_{j2}\right)$ through CIE-1931 XYZ values as described in 3.1. The colour difference between β_i and β_j under light source 1 will then be equal to ΔE^*_{ab1} in equation (9), and the colour difference between β_i and β_j under light source 2 will be equal to ΔE^*_{ab2} in equation (10). By comparing the differences with those for the human visual system, it becomes possible to evaluate a sensor's resemblance to human eyes.

$$\Delta E^*_{ab1} = \sqrt{\left(L^*_{i1} - L^*_{j1}\right)^2 + \left(a^*_{i1} - a^*_{j1}\right)^2 + \left(b^*_{i1} - b^*_{j1}\right)^2} \tag{9}$$

$$\Delta E^*_{ab2} = \sqrt{\left(L^*_{i2} - L^*_{j2}\right)^2 + \left(a^*_{i2} - a^*_{j2}\right)^2 + \left(b^*_{i2} - b^*_{j2}\right)^2} \tag{10}$$

4 Permissions

The primary intent of this Technical Report is to provide data in digital form for the use of individuals and organizations evaluating imaging systems. Therefore, the following is permitted:

a) The data files included as part of the Technical Report may be freely copied and used within the organization purchasing a copy of this Technical Report from ISO, or an authorized reseller of ISO documents. The data files (in either their original or any reformatted version) may not be distributed to any other individual or organization.

b) Where this database is used for research, development, or evaluation, any publication reporting such work shall identify this Technical Report as the source.

Annex A
(informative)

Spectral reflectance and transmittance source data

A.1 General

The spectral reflectance and transmittance source data collected as part of this project, and used to create the typical and difference set samples, are included as electronic attachments to this Technical Report. These data are classified into two groups, original data and interpolated data. The directory structure that contains these files is shown in Figure A.1.

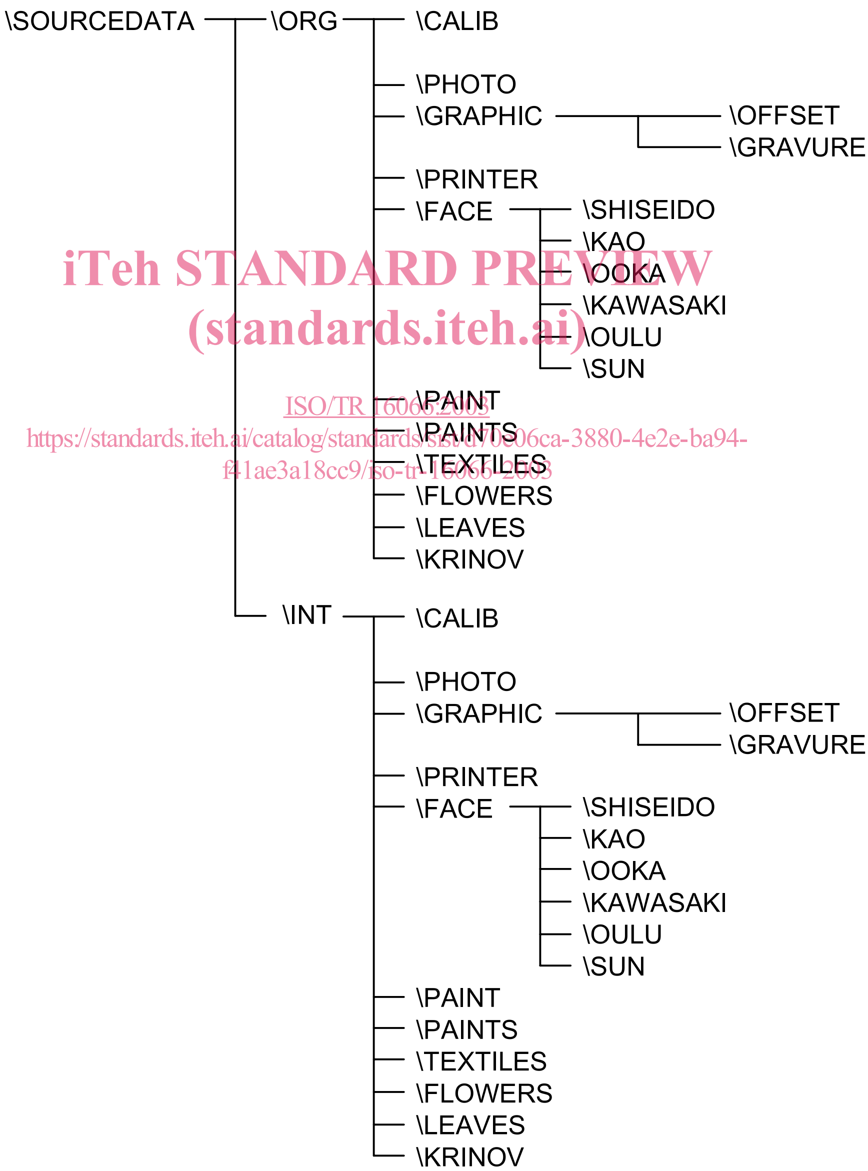

Figure A.1 — Directory structure of source data