



# The Printer Working Group (PWG) RepertoireSupported Element

February 1, 2004  
Candidate Standard 5101.2

## *Authors:*

Elliott Bradshaw, Zoran Imaging Division  
Ira McDonald, High North

Copyright 2004 Printer Working Group, All Rights Reserved.  
XHTML is a trademark of the World Wide Web Consortium.

An electronic version of this document is available online at:

<ftp://ftp.pwg.org/pub/pwg/standards/pwg5101.2.pdf>

## **Abstract**

In traditional printing environments, clients rely on font downloads when they are not sure a given character is embedded in the printer. As printing moves to small clients, downloading may not be an option and clients have a need to know what characters are available in a given device.

There are many published named character repertoires, and a small client will not know about them all.

To improve operability, this document defines semantics and naming conventions, to allow a printer to advertise what repertoires it supports.

The primary target of this document is printing using document formats based on XML or HTML (for example, XHTML-Print). It will be less applicable to traditional PDLs (PCL, PostScript, etc.) because they tend to have very format-specific mechanisms for managing character repertoires.

## **Notices**

**Copyright (C) 2004, IEEE ISTO. All rights reserved.**

This document may be copied and furnished to others, and derivative works that comment on, or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice, this paragraph and the title of the Document as referenced below are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the IEEE-ISTO and the Printer Working Group, a program of the IEEE-ISTO.

## RepertoireSupported Element

Title: The Printer Working RepertoireSupported Element

The IEEE–ISTO and the Printer Working Group DISCLAIM ANY AND ALL WARRANTIES, WHETHER EXPRESS OR IMPLIED INCLUDING (WITHOUT LIMITATION) ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

The Printer Working Group, a program of the IEEE–ISTO, reserves the right to make changes to the document without further notice. The document may be updated, replaced or made obsolete by other documents at any time.

The IEEE–ISTO takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights.

The IEEE–ISTO invites any interested party to bring to its attention any copyrights, patents, or patent applications, or other proprietary rights which may cover technology that may be required to implement the contents of this document. The IEEE–ISTO and its programs shall not be responsible for identifying patents for which a license may be required by a document and/or IEEE–ISTO Industry Group Standard or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention. Inquiries may be submitted to the IEEE–ISTO by e–mail at:

ieee–isto@ieee.org

The Printer Working Group acknowledges that the IEEE–ISTO (acting itself or through its designees) is, and shall at all times, be the sole entity that may authorize the use of certification marks, trademarks, or other special designations to indicate compliance with these materials.

Use of this document is wholly voluntary. The existence of this document does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to its scope.

All sections of this document are normative unless noted as informative.

Public discussion of PWG Character Repertoires takes place on the mailing list: [cr@pwg.org](mailto:cr@pwg.org). To subscribe send an email to [majordomo@pwg.org](mailto:majordomo@pwg.org) with the words *subscribe cr* in the body. You must be subscribed to the mailing list to post there. Please report errors in this document to the editor listed above or on the mailing list.

A list of current PWG Standards and other technical documents can be found at <http://www.pwg.org/standards.html>.

## 1. Terminology

We use the term **charset** as defined in [\[RFC2978\]](#), which says in part:

The term "charset" is used here to refer to a method of converting a sequence of octets into a sequence of characters.

We define the term **character repertoire** as a named subset of the characters defined in a given charset

## RepertoireSupported Element

standard (e.g., Unicode/4.0) that are supported for output rendering of document data. A repertoire, while defined in terms of one charset, may be used in the context of another charset (e.g., the value of "document-charset" in the the IPP Document object) through suitable mapping. For example, the repertoire "ISO 8859-7" may be used in a Unicode context, in which case it names the set of Unicode characters mapping to the underlying characters in ISO 8859-7.

The keywords "MUST", "SHALL", "MUST NOT", "SHALL NOT", "REQUIRED", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" when used in this document are to be interpreted as described in RFC 2119 [RFC2119]. However, for readability, these words do not appear in all uppercase letters in this specification.

## 2. Overview

In a bidirectional printing environment, a client device exchanges information with a printer. The client may be a traditional Windows PC or may be a lighter weight device, such as a:

- PDA
- Set-top box
- Cell phone

A client uses some transport mechanism (outside the scope of this specification) to obtain from a particular printer the supported values of charset and character repertoire. The present specification describes a mechanism for the client to determine what characters can be printed by the printer, using the supported values of charset and repertoire as supplied by the printer.

A data element for supported charset values is already described elsewhere by the Semantic Model element "DocumentCharsetSupported". The present specification describes an additional element for supported repertoires, called "RepertoireSupported".

For a given supported charset, the client can determine which characters are supported by the printer in the following way. For each character in the charset, if it is referenced in one or more supported repertoires, then that character is supported for printing. If it is not referenced in any of the supported repertoires, the character is not supported for printing, unless it is mandated by some feature of a document formatting language (see below).

Because each repertoire is defined with some particular encoding, it may be necessary to map repertoire coding values into corresponding coding values in the chosen charset when doing this calculation.

Some document formats allow for escape sequences or other higher-level syntax to access characters using numeric values; for example HTML uses the "&#176;" syntax to access a degree character. In such a case, the character is supported if it is referenced in one or more supported repertoires. Again, mapping may be required.

Some document formats allow for named characters; for example XHTML-Print uses the "&deg;" syntax to access a degree character. If a format requires support for a particular named character, the printer must support it regardless of what repertoires it advertises.

The data element "RepertoireSupported" is intended to be incorporated into higher level description schemes, such as the PWG Semantic Model [PWG-SM], as well as protocols based on those schemes.

## RepertoireSupported Element

Inside the scope of this document are:

1. Syntactic conventions for advertisement of character repertoires defined elsewhere.
2. Rules for a conforming printer to use when advertising supported repertoires.

A companion Best Practices document deals with recommended methods of implementation to improve interoperability between clients and printers.

Some areas outside the scope of either of these documents are:

1. Character encoding. It is assumed that the client and printer have some other way of agreeing on encoding.
2. Mapping into and out of Unicode. It is assumed that for any repertoire defined in a different encoding (e.g. ISO-Latin-xxx???, Shift-JIS), the implementer can provide a suitable mapping into Unicode.
3. Font downloading.
4. Adaptation to mature PDLs such as PostScript and PCL. These provide rich, alternate schemes for managing repertoires (including download), and it is not apparent how they would use the mechanisms in this document.
5. Ability to advertise individual characters. Our view is that this will add a great deal of data with little real-world benefit.
6. Query mechanisms. Separately, a protocol could define a query mechanism using this data format.

### 3. The Semantic Element "RepertoireSupported"

[PWG-SM] defines semantic elements for a printer to use in advertising its capabilities (among other things). We use the Model to let a printer advertise its supported repertoires; the union of all characters in all advertised repertoires tells the client what characters it may safely use. (Note that a printer is free to implement additional characters beyond those listed in the supported repertoires.)

#### 3.1. Syntax

The value of the element "RepertoireSupported" is made up of one more character repertoire names. These names are constructed from lists maintained elsewhere; a special prefix serves to identify the underlying source and to create a unique string value.

Names taken from elsewhere are mapped according to these rules:

1. Uppercase alpha characters are mapped to lower case.
2. Characters that are alphanumeric, "-", ".", or "\_" are preserved.
3. Other characters (including spaces) are converted to hyphen "-" characters.

Names are constructed as follows:

| Source                                              | Form of each value  | Example                                                               |
|-----------------------------------------------------|---------------------|-----------------------------------------------------------------------|
| IANA charset registry as defined in [IANA-Charsets] | <i>iana_name</i>    | iana_iso_8859-1<br>(based on IANA "ISO_8859-1")                       |
| Unicode code chart as defined in [Unicode-Charts]   | <i>unicode_name</i> | unicode_latin-1-supplement<br>(based on Unicode "Latin-1 Supplement") |

## RepertoireSupported Element

|                 |                           |                            |
|-----------------|---------------------------|----------------------------|
| Vendor specific | <i>vendor_vendor_name</i> | <i>vendor_zoran_floral</i> |
|-----------------|---------------------------|----------------------------|

When referring to IANA charsets, only these names are legal:

- Those marked "Name:" in [\[IANA-Charsets\]](#)
- Those marked "preferred Mime name" in [\[IANA-Charsets\]](#)
- Those marked "preferred Mime name" in another RFC

Other aliases are not legal, even if listed in [\[IANA-Charsets\]](#).

Note that IANA charsets are used to indicate character repertoires, because these are well defined and widely used. A charset provides both a list of characters, as well as encodings for each characters. However, a character repertoire is an abstract list of characters, which can be encoded in any number of ways. Therefore, when a charset is used to indicate a character repertoire, the specific encoding for that charset is irrelevant.

### 3.2. Meaning

By naming one or more supported repertoires, a complying printer guarantees support as follows:

1. Each character in each repertoire can be rendered in a recognizable way, regardless of currently selected font. However, renderings in different fonts need not be distinct. A common approach is for the printer to implement a system default font with all advertised characters, and to implement a fall-through mechanism that will render a character from the default font if it is not available in a currently selected font.
2. When an advertised repertoire is encoded differently from an advertised charset, the printer supports those characters in the advertised charset which map to the advertised repertoire.
3. When the printer advertises supported charsets, there may be characters in advertised repertoires that are not in any advertised charset. Nonetheless the printer must support them, because a client may ask for such characters via a higher-level mechanism (e.g., character entities in HTML).

In addition to characters in advertised repertoires, a printer may support additional characters, which may or may not be available in all fonts.

A client references characters in whatever encoding is present, without reference to a particular repertoire. In other words, repertoires are (possibly overlapping) sets of characters, but a repertoire is not needed to reference a character. Therefore, there are no semantic elements for default, current, or actual repertoire values.

### 3.3. Combined Forms and Other Character Conversions

Unicode and many other charsets define a variety of ways to convert from one character sequence to another within the same charset. A common example is to use a single accented character (a "combined form") to represent what could also be specified as a two-character sequence of base character plus separate accent (a "decomposed form").

This document takes no position with respect to such visually equivalent character sequences. A client must not make any assumptions about a printer's support for such character sequence conversions. If a printer advertises support for a base character and an accent, then that printer must also specifically advertise the

combined form, if it is also supported.

## 4. Acknowledgements (informative)

This document was prepared with input and assistance from:

- Rick Seeler, Adobe
- Rod Acosta, Agfa Monotype
- Lee Farrell, Canon
- Jun Fujisawa, Canon
- Michael Sweet, Easy Software
- Michael Wu, Heidelberg
- Jim Bigelow, HP
- Harry Lewis, IBM
- Paul Tykodi, Intermate US
- Mark Robb, Lexmark
- Don Wright, Lexmark
- Don Levinstone, Motorola
- Tom Hastings, Xerox
- Peter Zehler, Xerox

## A. References

### [IANA–Charsets]

Available online at <http://www.iana.org/assignments/character-sets>.

### [PWG–SM]

IEEE–ISTO 5105.1–2004 The Printer Working Group Semantic Model, January 20, 2004, P. Zehler. <ftp://ftp.pwg.org/pub/pwg/standards/pwg5105.1.pdf>, Candidate Standard.

### [RFC2119]

"RFC2119 – Key words for use in RFCs to Indicate Requirement Levels", S. Bradner. Available online at <http://www.ietf.org/rfc/rfc2119>.

### [RFC2978]

"RFC2978 – IANA Charset Registration Procedures", N. Freed. Available online at <http://www.ietf.org/rfc/rfc2978.txt>

### [Unicode–Charts]

Unicode code charts, version 4.0. Beta code charts available online at <http://www.unicode.org/charts/u40-beta.html>.

## B. Bindings to IPP (Normative)

```
repertoire-supported (1setOf (keyword | name))
```

This REQUIRED IPP Printer Description attribute identifies some or all of the character repertoires that the IPP Printer object and contained IPP Job objects support for rendering of document data content.

The ABNF [RFC2234] for legal values of "repertoire-supported" is:

```
repertoire = rep-prefix "_" rep-name  
rep-prefix = "unicode" / ; from Code Chart titles
```

## 4. Acknowledgements (informative)

## RepertoireSupported Element

```

                                ; of Unicode/4.0 char database
                                ; from Name or Alias fields in
                                ; IANA Charset Registry
                                ; from vendor-specific
                                ; repertoire names
"iana" /
"vendor"
rep-name  = rep-alpha *(rep-char)
rep-char  = rep-alpha / rep-digit /      ; alphanumeric or
          "-" / "." / "_"              ; limited punctuation chars
rep-alpha = %x61-7A                    ; lowercase a-z
rep-digit = %x30-39                    ; decimal 0-9
```