ISO 28500:2017-08 (E)

Information and documentation - WARC file format

Contents

Page

Foreword v

Introduction vi

1 Scope 1

2 Normative references 1

3 Terms, definitions and abbreviated terms 2

4 File and record model 3

5 Named fields 5
5.1 General 5
5.2 WARC-Record-ID (mandatory) 5
5.3 Content-Length (mandatory) 5
5.4 WARC-Date (mandatory) 6
5.5 WARC-Type (mandatory) 6
5.6 Content-Type 6
5.7 WARC-Concurrent-To 7
5.8 WARC-Block-Digest 7
5.9 WARC-Payload-Digest 7
5.10 WARC-IP-Address 8
5.11 WARC-Refers-To 8
5.12 WARC-Refers-To-Target-URI 8
5.13 WARC-Refers-To-Date 8
5.14 WARC-Target-URI 9
5.15 WARC-Truncated 9
5.16 WARC-Warcinfo-ID 9
5.17 WARC-Filename 9
5.18 WARC-Profile 10
5.19 WARC-Identified-Payload-Type 10
5.20 WARC-Segment-Number 10
5.21 WARC-Segment-Origin-ID 10
5.22 WARC-Segment-Total-Length 10

6 WARC record types 11
6.1 General 11
6.2 `warcinfo' 11
6.3 `response' 11
6.3.1 General 11
6.3.2 `http' and `https' schemes 12
6.3.3 Other URI schemes 12
6.4 `resource' 12
6.4.1 General 12
6.4.2 `http' and `https' schemes 12
6.4.3 `ftp' scheme 12
6.4.4 `dns' scheme 13
6.4.5 Other URI schemes 13
6.5 `request' 13
6.5.1 General 13
6.5.2 `http' and `https' schemes 13

6.5.3 Other URI schemes 13
6.6 `metadata' 13
6.7 `revisit' 14
6.7.1 General 14
6.7.2 Profile: Identical Payload Digest 14
6.7.3 Profile: Server Not Modified 15
6.7.4 Other profiles 15
6.8 `conversion' 15
6.9 `continuation' 16

7 Record segmentation 16

8 WARC file name, size and compression 16

Annex A (informative) Use cases for writing WARC records 18

Annex B (informative) Examples of WARC records 21

Annex C (informative) WARC file size and name recommendations 24

Annex D (informative) Compression recommendations 25

Bibliography 26