ISO 28500:2009-05 (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
3.1 Terms and definitions 2
3.2 Abbreviated terms 2

4 File and record model 3

5 Named fields 5
5.1 General 5
5.2 WARC-Record-ID (mandatory) 6
5.3 Content-Length (mandatory) 6
5.4 WARC-Date (mandatory) 6
5.5 WARC-Type (mandatory) 6
5.6 Content-Type 7
5.7 WARC-Concurrent-To 7
5.8 WARC-Block-Digest 8
5.9 WARC-Payload-Digest 8
5.10 WARC-IP-Address 8
5.11 WARC-Refers-To 9
5.12 WARC-Target-URI 9
5.13 WARC-Truncated 9
5.14 WARC-Warcinfo-ID 10
5.15 WARC-Filename 10
5.16 WARC-Profile 10
5.17 WARC-Identified-Payload-Type 10
5.18 WARC-Segment-Number 10
5.19 WARC-Segment-Origin-ID 11
5.20 WARC-Segment-Total-Length 11

6 WARC record types 11
6.1 General 11
6.2 'warcinfo' 11
6.3 'response' 12
6.4 'resource' 13
6.5 'request' 13
6.6 'metadata' 14
6.7 'revisit' 15
6.8 'conversion' 16
6.9 'continuation' 16

7 Record segmentation 16

8 Registration of MIME media types application/warc and application/warc-fields 17
8.1 General 17
8.2 application/warc 17

8.3 application/warc-fields 18

9 WARC file name, size and compression 18

Annex A (informative) Use cases for writing WARC records 19

Annex B (informative) Examples of WARC records 22

Annex C (informative) WARC file size and name recommendations 26

Annex D (informative) Compression recommendations 27

Bibliography 28