

# ISO 28500:2009-05 (E)

## Information and documentation - WARC file format

---

| <b>Contents</b>    |  | <b>Page</b> |
|--------------------|--|-------------|
| Foreword .....     |  | v           |
| Introduction ..... |  | vi          |
| <b>1</b>           | <b>Scope .....</b>   | <b>1</b>    |
| <b>2</b>           | <b>Normative references .....</b>  | <b>1</b>    |
| <b>3</b>           | <b>Terms, definitions and abbreviated terms .....</b>                                      | <b>2</b>    |
| <b>3.1</b>         | <b>Terms and definitions .....</b>   | <b>2</b>    |
| <b>3.2</b>         | <b>Abbreviated terms .....</b>   | <b>2</b>    |
| <b>4</b>           | <b>File and record model .....</b>   | <b>3</b>    |
| <b>5</b>           | <b>Named fields .....</b>  | <b>5</b>    |
| <b>5.1</b>         | <b>General .....</b>   | <b>5</b>    |
| <b>5.2</b>         | <b>WARC-Record-ID (mandatory) .....</b>  | <b>6</b>    |
| <b>5.3</b>         | <b>Content-Length (mandatory) .....</b>  | <b>6</b>    |
| <b>5.4</b>         | <b>WARC-Date (mandatory) .....</b>   | <b>6</b>    |
| <b>5.5</b>         | <b>WARC-Type (mandatory) .....</b>   | <b>6</b>    |
| <b>5.6</b>         | <b>Content-Type .....</b>  | <b>7</b>    |
| <b>5.7</b>         | <b>WARC-Concurrent-To .....</b>  | <b>7</b>    |
| <b>5.8</b>         | <b>WARC-Block-Digest .....</b>   | <b>8</b>    |
| <b>5.9</b>         | <b>WARC-Payload-Digest .....</b>   | <b>8</b>    |
| <b>5.10</b>        | <b>WARC-IP-Address .....</b>   | <b>8</b>    |
| <b>5.11</b>        | <b>WARC-Refers-To .....</b>  | <b>9</b>    |
| <b>5.12</b>        | <b>WARC-Target-URI .....</b>   | <b>9</b>    |
| <b>5.13</b>        | <b>WARC-Truncated .....</b>  | <b>9</b>    |
| <b>5.14</b>        | <b>WARC-Warcinfo-ID .....</b>  | <b>10</b>   |
| <b>5.15</b>        | <b>WARC-Filename .....</b>   | <b>10</b>   |
| <b>5.16</b>        | <b>WARC-Profile .....</b>  | <b>10</b>   |
| <b>5.17</b>        | <b>WARC-Identified-Payload-Type .....</b>  | <b>10</b>   |
| <b>5.18</b>        | <b>WARC-Segment-Number .....</b>   | <b>10</b>   |
| <b>5.19</b>        | <b>WARC-Segment-Origin-ID .....</b>  | <b>11</b>   |
| <b>5.20</b>        | <b>WARC-Segment-Total-Length .....</b>   | <b>11</b>   |
| <b>6</b>           | <b>WARC record types .....</b>   | <b>11</b>   |
| <b>6.1</b>         | <b>General .....</b>   | <b>11</b>   |
| <b>6.2</b>         | <b>'warcinfo' .....</b>  | <b>11</b>   |
| <b>6.3</b>         | <b>'response' .....</b>  | <b>12</b>   |
| <b>6.4</b>         | <b>'resource' .....</b>  | <b>13</b>   |
| <b>6.5</b>         | <b>'request' .....</b>   | <b>13</b>   |
| <b>6.6</b>         | <b>'metadata' .....</b>  | <b>14</b>   |
| <b>6.7</b>         | <b>'revisit' .....</b>   | <b>15</b>   |
| <b>6.8</b>         | <b>'conversion' .....</b>  | <b>16</b>   |
| <b>6.9</b>         | <b>'continuation' .....</b>  | <b>16</b>   |
| <b>7</b>           | <b>Record segmentation .....</b>   | <b>16</b>   |
| <b>8</b>           | <b>Registration of MIME media types application/warc and application/warc-fields .....</b> | <b>17</b>   |
| <b>8.1</b>         | <b>General .....</b>   | <b>17</b>   |
| <b>8.2</b>         | <b>application/warc .....</b>  | <b>17</b>   |

|     |   |    |
|-----|---|----|
| 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 |