

# Caveat user

Known side-effects and pitfalls from the current reading/validating strategy

## Side-effects:

### **The payload is inaccessible when the "Content-Length" is absent or invalid.**

For GZip'ed records this is not a big problem since we know the record ends when the GZip entry ends.

For uncompressed records the payload input stream would have to look ahead for a valid ARC/WARC header at which point the payload stream should be closed and the bytes read beyond that pushed back onto the internal streams.

In these cases the payload is inaccessible and errors are reported. ([JIRA issue](#))

### **Warc-Payload-Digest header is computed only on defined record payloads**

If the payload can not be identified it will not be recognized as WARC payload and treated as a normal WARC block. This makes it a requirement for the WARC parser to identify and always parse the http response and not make it optional.

Currently the payload processor always tries to identify the payload content. Since the identifiable payload header could be of use it is available through the API.