

# Media Types and Data Definition Languages

Discussion point

Carsten Bormann

CoRE@IETF 107+, 2020-04-29, in the cloud

# Media types and text-based data: A tacit rule (since ~RFC 733, 1977)

- If a media type uses text-based data (ASCII, UTF-8), and
- There is no structure governing the text (e.g., JSON), and
- No other data definition language is in use (e.g., YANG),
- → We should use ABNF (RFC 5234+7405) to provide a machine-readable definition of the text-based data in the media type instances.

# Media types and structured data:

## A tacit rule

- If a media type uses structured data (CBOR, JSON), and
- No other data definition language is in use (e.g., YANG),
- → We should use CDDL (RFC 8610) to provide a machine-readable definition of the structured data in the media type instances.

# Why use a machine-readable definition language/format?

- English language definitions are ambiguous and lengthy
- With machine-readable definitions, CI (continuous integration) can check examples right from the –00 draft,
  - and readers can check their understanding, too
- Machine-readable definitions can be used in implementations for
  - Generating parsers/ingestion code
  - Validation of instances
- Fewer doubtful cases in interop testing, more interoperability

# Why **not** use a machine-readable definition language/format?

- Definition languages need to be learned — threshold effect
  - Mitigated by focusing on very few of them
- Definition languages can lead to tolerance for excessive complexity
  - The well-known “ASN.1 effect”
  - This can only be mitigated by discipline
  - Example (RFC 8710): Two lines of CDDL suffice
  - Counterexample: GSM MAP (TS 09.02), ~ 500 pages of ASN.1

# What do we get from a data definition?

- Validation: Defined **shape** of data expected
  - What must be there, what is optional
  - Valid ranges for variables
  - Identify extension points (advanced feature)
- Augmentation: Expose anchor points for adding semantics
  - Implicit (by just naming components or string-valued labels), (semantics then usually added in English)
  - Explicit (e.g., pointing to RDF universe)

# Assuming Tacit Consensus

Until here...

Entering wild speculation now.

# Background: The missed-level mistake

- RFC 5988 defines a structured object, the **Web Link**
- Web Links are represented in text, so 5988 uses ABNF to define the text form (including quoting etc.)

```
Link = "Link" ":" #link-value
```

```
link-value = "<" URI-Reference ">" *( ";" link-param )
```

- Web Links also have structural aspects, and 5988 tries to use the same ABNF to define that next-higher level as well:

```
      | ( "media" "=" ( MediaDesc | ( <"> MediaDesc <"> ) ) )  
      | ( "title" "=" quoted-string )
```

- RFC 6690 (CoRE link format) inherits this mistake
- RFC 8288 fixes this, provides ABNF for **higher-level** data (quote-stripped) only



# POCI/POJI vs. CoRAL

- CDDL addresses “plain old CBOR”, “plain old JSON” items
- What about CoRAL-shaped media types?
- Is “it’s CoRAL” enough of a data definition?
  - True statement
  - But should we also talk about the next higher level of structure?
- Again, need for
  - Minimum expected content, ranges/types etc. (validation)
  - Additional interpretation? (augmentation)

# Example: Problem Details Clone Draft

- RFC 7807 → draft-fossati-core-coap-problem (in WGA call)
- Pretty straightforward as a POCO
  - Some weirdness due to ns/type naming (replaced by URI in CoRAL), separate issue
- Similar considerations when represented as CoRAL objects
  - Validation: We do want the type relation to be there; rest optional
  - Augmentation is implicit by the link relation types
- Do we need a CoDDL for CoRAL?

# Embedding CDDL/CoDDL in CoRAL?

- Interaction affordance (link/form) expressed in CoRAL could directly provide data definition information (send/expect content-type X → send/expect this DD)
- Could occur anywhere where a media type goes now (or even both together? application/cbor + some CDDL?)