API Guidelines

Table of Contents

Zalando RESTful API and Event Guidelines

GitHub Repository as part of Zalando SE Opensource

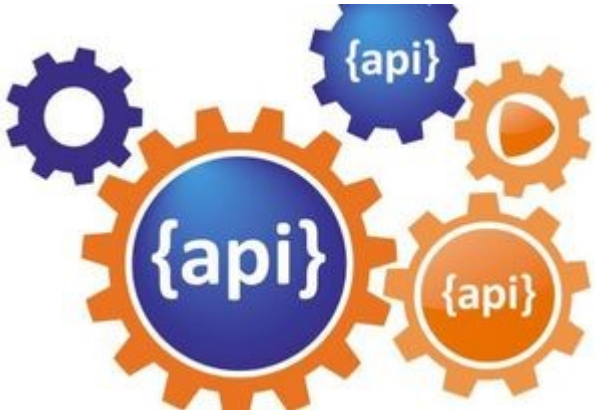

Table of Contents

- **MUST** use standard data formats
- **MUST** define a format for number and integer types
- **MUST** use standard formats for date and time properties
- **SHOULD** select appropriate one of date or date-time format
- **SHOULD** use standard formats for time duration and interval properties
- **MUST** use standard formats for country, language and currency properties
- **SHOULD** use content negotiation, if clients may choose from different resource representations
- **SHOULD** only use UUIDs if necessary

7. REST Basics - URLs
   - **SHOULD** not use /api as base path
   - **MUST** pluralize resource names
   - **MUST** use URL-friendly resource identifiers
   - **MUST** use kebab-case for path segments
   - **MUST** use normalized paths without empty path segments and trailing slashes
   - **MUST** keep URLs verb-free
   - **MUST** avoid actions — think about resources
   - **SHOULD** define *useful* resources
   - **MUST** use domain-specific resource names
   - **SHOULD** model complete business processes
   - **MUST** identify resources and sub-resources via path segments
   - **MAY** expose compound keys as resource identifiers
   - **MAY** consider using (non-) nested URLs
   - **SHOULD** limit number of resource types
   - **SHOULD** limit number of sub-resource levels
   - **MUST** use snake_case (never camelCase) for query parameters
   - **MUST** stick to conventional query parameters
8. REST Basics - JSON payload
   - **MUST** use JSON as payload data interchange format
   - **SHOULD** design single resource schema for reading and writing
   - **SHOULD** be aware of services not fully supporting JSON/unicode
   - **MAY** pass non-JSON media types using data specific standard formats
   - **SHOULD** use standard media types
   - **SHOULD** pluralize array names
   - **MUST** property names must be snake_case (and never camelCase)
   - **SHOULD** declare enum values using UPPER_SNAKE_CASE string

- **SHOULD** use naming convention for date/time properties
- **SHOULD** define maps using `additionalProperties`
- **MUST** use same semantics for `null` and absent properties
- **MUST** not use `null` for boolean properties
- **SHOULD** not use `null` for empty arrays
- **MUST** use common field names and semantics
- **MUST** use the common address fields
- **MUST** use the common money object

9. REST Basics - HTTP requests
   - **MUST** use HTTP methods correctly
   - **MUST** fulfill common method properties
   - **SHOULD** consider to design `POST` and `PATCH` idempotent
   - **SHOULD** use secondary key for idempotent `POST` design
   - **MAY** support asynchronous request processing
   - **MUST** define collection format of header and query parameters
   - **SHOULD** design simple query languages using query parameters
   - **SHOULD** design complex query languages using JSON
   - **MUST** document implicit response filtering
10. REST Basics - HTTP status codes
    - **MUST** use official HTTP status codes
    - **MUST** specify success and error responses
    - **SHOULD** only use most common HTTP status codes
    - **MUST** use most specific HTTP status codes
    - **MUST** use code 207 for batch or bulk requests
    - **MUST** use code 429 with headers for rate limits
    - **MUST** support problem JSON
    - **MUST** not expose stack traces
    - **SHOULD** not use redirection codes
11. REST Basics - HTTP headers

    Using Standard Header definitions
    - **MAY** use standard headers
    - **SHOULD** use kebab-case with uppercase separate words for HTTP headers
    - **MUST** use `Content-*` headers correctly
    - **SHOULD** use `Location` header instead of `Content-Location` header
    - **MAY** use `Content-Location` header
    - **MAY** consider to support `Prefer` header to handle processing preferences

- **MAY** consider to support `ETag` together with `If-Match` / `If-None-Match` header
- **MAY** consider to support `Idempotency-Key` header
- **SHOULD** use only the specified proprietary Zalando headers
- **MUST** propagate proprietary headers
- **MUST** support `X-Flow-ID`

12. REST Design - Hypermedia
    - **MUST** use REST maturity level 2
    - **MAY** use REST maturity level 3 - HATEOAS
    - **MUST** use common hypertext controls
    - **SHOULD** use simple hypertext controls for pagination and self-references
    - **MUST** use full, absolute URI for resource identification
    - **MUST** not use link headers with JSON entities
13. REST Design - Performance
    - **SHOULD** reduce bandwidth needs and improve responsiveness
    - **SHOULD** use `gzip` compression
    - **SHOULD** support partial responses via filtering
    - **SHOULD** allow optional embedding of sub-resources
    - **MUST** document cacheable `GET` , `HEAD` , and `POST` endpoints
14. REST Design - Pagination
    - **MUST** support pagination
    - **SHOULD** prefer cursor-based pagination, avoid offset-based pagination
    - **SHOULD** use pagination response page object
    - **SHOULD** use pagination links
    - **SHOULD** avoid a total result count
15. REST Design - Compatibility
    - **MUST** not break backward compatibility
    - **SHOULD** prefer compatible extensions
    - **SHOULD** design APIs conservatively
    - **MUST** prepare clients to accept compatible API extensions
    - **MUST** treat OpenAPI specification as open for extension by default
    - **SHOULD** avoid versioning
    - **MUST** use media type versioning
    - **MUST** not use URL versioning
    - **MUST** always return JSON objects as top-level data structures
    - **SHOULD** use open-ended list of values (`x-extensible-enum`) for enumeration types
16. REST Design - Deprecation

MUST reflect deprecation in API specifications

MUST obtain approval of clients before API shut down

MUST collect external partner consent on deprecation time span

MUST monitor usage of deprecated API scheduled for sunset

SHOULD add `Deprecation` and `Sunset` header to responses

SHOULD add monitoring for `Deprecation` and `Sunset` header

MUST not start using deprecated APIs

17. REST Operation

MUST publish OpenAPI specification for non-component-internal APIs

SHOULD monitor API usage

18. EVENT Basics - Event Types

MUST define events compliant with overall API guidelines

MUST treat events as part of the service interface

MUST make event schema available for review

MUST specify and register events as event types

MUST follow naming convention for event type names

MUST indicate ownership of event types

MUST carefully define the compatibility mode

MUST ensure event schema conforms to OpenAPI schema object

SHOULD avoid `additionalProperties` in event type schemas

MUST use semantic versioning of event type schemas

19. EVENT Basics - Event Categories

MUST ensure that events conform to an event category

MUST provide mandatory event metadata

MUST provide unique event identifiers

MUST use general events to signal steps in business processes

SHOULD provide explicit event ordering for general events

MUST use data change events to signal mutations

MUST provide explicit event ordering for data change events

SHOULD use the hash partition strategy for data change events

20. EVENT Design

SHOULD avoid writing sensitive data to events

MUST be robust against duplicates when consuming events

SHOULD design for idempotent out-of-order processing

MUST ensure that events define useful business resources

SHOULD ensure that data change events match the APIs resources

1. Introduction

Zalando's software architecture centers around decoupled microservices that provide functionality via RESTful APIs with a JSON payload. Small engineering teams own, deploy and operate these microservices in their AWS (team) accounts. Our APIs express most purely what our systems do, and are therefore highly valuable business assets. Designing high-quality, long-lasting APIs has become even more critical for us since we started developing our new open platform strategy, which transforms Zalando from an online shop into an expansive fashion platform. Our strategy emphasizes developing lots of public APIs for our external business partners to use via third-party applications.

With this in mind, we've adopted "API First" as one of our key engineering principles. Microservices development begins with API definition outside the code and ideally involves ample peer-review feedback to achieve high-quality APIs. API First encompasses a set of quality-related standards and fosters a peer review culture including a lightweight review procedure. We encourage our teams to follow them to ensure that our APIs:

- are easy to understand and learn

- are general and abstracted from specific implementation and use cases
- are robust and easy to use
- have a common look and feel
- follow a consistent RESTful style and syntax
- are consistent with other teams' APIs and our global architecture

Ideally, all Zalando APIs will look like the same author created them.

Conventions used in these guidelines

The requirement level keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" used in this document (case insensitive) are to be interpreted as described in RFC 2119.

Zalando specific information

The purpose of our "RESTful API guidelines" is to define standards to successfully establish "consistent API look and feel" quality. The API Guild (internal_link) drafted and owns this document. Teams are responsible to fulfill these guidelines during API development and are encouraged to contribute to guideline evolution via pull requests.

These guidelines will, to some extent, remain work in progress as our work evolves, but teams can confidently follow and trust them.

In case guidelines are changing, following rules apply:

- existing APIs don't have to be changed, but we recommend it
- clients of existing APIs have to cope with these APIs based on outdated rules
- new APIs have to respect the current guidelines

Furthermore you should keep in mind that once an API becomes public externally available, it has to be re-reviewed and changed according to current guidelines - for sake of overall consistency.

---

2. Principles

API design principles

Comparing SOA web service interfacing style of SOAP vs. REST, the former tend to be centered around operations that are usually use-case specific and specialized. In contrast, REST is centered around business (data) entities exposed as resources that are identified via URIs and can be manipulated via standardized CRUD-like methods using different representations, and hypermedia. RESTful APIs tend to be less use-case specific and come with less rigid client / server coupling and are more suitable for an ecosystem of (core) services providing a platform of APIs to build diverse new business services. We apply the RESTful web service principles to all kind of application (micro-) service components, independently from whether they provide functionality via the internet or intranet.

- We prefer REST-based APIs with JSON payloads
- We prefer systems to be truly RESTful [1]

An important principle for API design and usage is Postel's Law, aka The Robustness Principle (see also RFC 1122):

- Be liberal in what you accept, be conservative in what you send

Readings: Some interesting reads on the RESTful API design style and service architecture:

- Article: REST API Design - Resource Modeling
- Article: Richardson Maturity Model — Steps toward the glory of REST
- Book: Irresistible APIs: Designing web APIs that developers will love
- Book: REST in Practice: Hypermedia and Systems Architecture
- Book: Build APIs You Won't Hate
- Fielding Dissertation: Architectural Styles and the Design of Network-Based Software Architectures

API as a product

As mentioned above, Zalando is transforming from an online shop into an expansive fashion platform comprising a rich set of products following a Software as a Platform (SaaP) model for our business partners. As a company we want to deliver products to our (internal and external) customers which can be consumed like a service.

Platform products provide their functionality via (public) APIs; hence, the design of our APIs should be based on the API as a Product principle:

- Treat your API as product and act like a product owner
- Put yourself into the place of your customers; be an advocate for their needs
- Emphasize simplicity, comprehensibility, and usability of APIs to make them irresistible for client engineers
- Actively improve and maintain API consistency over the long term
- Make use of customer feedback and provide service level support

Embracing 'API as a Product' facilitates a service ecosystem, which can be evolved more easily and used to experiment quickly with new business ideas by recombining core capabilities. It makes the difference between agile, innovative product service business built on a platform of APIs and ordinary enterprise integration business where APIs are provided as "appendix" of existing products to support system integration and optimised for local server-side realization.

Understand the concrete use cases of your customers and carefully check the trade-offs of your API design variants with a product mindset. Avoid short-term implementation optimizations at the expense of unnecessary client side obligations, and have a high attention on API quality and client developer experience.

API as a Product is closely related to our API First principle (see next chapter) which is more focused on how we engineer high quality APIs.

API first

API First is one of our engineering and architecture principles. In a nutshell API First requires two aspects:

- define APIs first, before coding its implementation, using a standard specification language
- get early review feedback from peers and client developers

By defining APIs outside the code, we want to facilitate early review feedback and also a development discipline that focus service interface design on…

- profound understanding of the domain and required functionality
- generalized business entities / resources, i.e. avoidance of use case specific APIs

- clear separation of WHAT vs. HOW concerns, i.e. abstraction from implementation aspects — APIs should be stable even if we replace complete service implementation including its underlying technology stack

Moreover, API definitions with standardized specification format also facilitate…

- single source of truth for the API specification; it is a crucial part of a contract between service provider and client users
- infrastructure tooling for API discovery, API GUIs, API documents, automated quality checks

Elements of API First are also this API Guidelines and a standardized API review process as to get early review feedback from peers and client developers. Peer review is important for us to get high quality APIs, to enable architectural and design alignment and to supported development of client applications decoupled from service provider engineering life cycle.

It is important to learn, that API First is **not in conflict with the agile development principles** that we love. Service applications should evolve incrementally — and so its APIs. Of course, our API specification will and should evolve iteratively in different cycles; however, each starting with draft status and *early* team and peer review feedback. API may change and profit from implementation concerns and automated testing feedback. API evolution during development life cycle may include breaking changes for not yet productive features and as long as we have aligned the changes with the clients. Hence, API First does *not* mean that you must have 100% domain and requirement understanding and can never produce code before you have defined the complete API and get it confirmed by peer review.

On the other hand, API First obviously is in conflict with the bad practice of publishing API definition and asking for peer review after the service integration or even the service productive operation has started. It is crucial to request and get early feedback — as early as possible, but not before the API changes are comprehensive with focus to the next evolution step and have a certain quality (including API Guideline compliance), already confirmed via team internal reviews.

---

3. General guidelines

The titles are marked with the corresponding labels: **MUST**, **SHOULD**, **MAY**.

MUST follow API first principle [100]

You must follow the API First Principle, more specifically:

- You must define APIs first, before coding its implementation, using OpenAPI as specification language
- You must design your APIs consistently with these guidelines; use our API Linter Service (internal_link) for automated rule checks.
- You must call for early review feedback from peers and client developers, and apply our lightweight API review process (internal_link) for all component external APIs, i.e. all apis with `x-api-audience =/= component-internal` (see **MUST** provide API audience).

MUST provide API specification using OpenAPI [101]

We use the OpenAPI specification as standard to define API specification files. API designers are required to provide the API specification using a single **self-contained YAML** file to improve readability. We encourage to use **OpenAPI 3.0** version, but still support **OpenAPI 2.0** (a.k.a. Swagger 2).

The API specification files should be subject to version control using a source code management system - best together with the implementing sources.

You must / should publish the component external / internal API specification with the deployment of the implementing service, and, hence, make it discoverable for the group via our API Portal (internal_link).

Hint: A good way to explore **OpenAPI 3.0/2.0** is to navigate through the OpenAPI specification mind map and use our Swagger Plugin for IntelliJ IDEA to create your first API. To explore and validate/evaluate existing APIs the Swagger Editor or our API Portal may be a good starting point.

Hint: We do not yet provide guidelines for GraphQL and focus on resource oriented HTTP/REST API style (and related tooling and infrastructure support). Following our Zalando Tech Radar (internal_link), we think that GraphQL has no major benefits, but a couple of downsides compared to REST as API technology for general purpose peer-to-peer microservice communication. However, GraphQL can provide a lot of value for specific target domain problems, especially backends for frontends (BFF) and mobile clients, where typically many (domain object) resources from different services are queried and multiple roundtrip overhead should be avoided due to (mobile or public) network constraints. Therefore we list both technologies on ADOPT, though GraphQL only supplements REST for the BFF-specific problem

domain.

SHOULD provide API user manual [102]

In addition to the API Specification, it is good practice to provide an API user manual to improve client developer experience, especially of engineers that are less experienced in using this API. A helpful API user manual typically describes the following API aspects:

- API scope, purpose, and use cases
- concrete examples of API usage
- edge cases, error situation details, and repair hints
- architecture context and major dependencies - including figures and sequence flows

The user manual must be published online, e.g. via our documentation hosting platform service, GHE pages, or specific team web servers. Please do not forget to include a link to the API user manual into the API specification using the `#/externalDocs/url` property.

MUST write APIs using U.S. English [103]

MUST only use durable and immutable remote references [234]

Normally, API specification files must be **self-contained**, i.e. files should not contain references to local or remote content, e.g. `../fragment.yaml#/element` or `$ref: 'https://github.com/zalando/zally/blob/master/server/src/main/resources/api/zally-api.yaml#/schemas/LintingRequest'` . The reason is, that the content referred to is *in general* **not durable** and **not immutable**. As a consequence, the semantic of an API may change in unexpected ways. (For example, the second link is already outdated due to code restructuring.)

However, you may use remote references to resources accessible by the following service URLs:

- `https://infrastructure-api-repository.zalandoapis.com/ (internal_link)` – used to refer to user-defined, immutable API specification revisions published via the internal API repository.
- `https://opensource.zalando.com/restful-api-guidelines/{model.yaml}` – used to refer to guideline-defined re-usable API fragments (see `{model.yaml}` files in restful-api-guidelines/models for details).

Hint: The formerly used remote references to the `Problem` API fragment (aliases `https://opensource.zalando.com/problem/` and `https://zalando.github.io/problem/`) are deprecated, but still supported for compatibility (**MUST** support problem JSON on how to replace).

As we control these URLs, we ensure that their content is **durable** and **immutable**. This allows to define API specifications by using fragments published via these sources, as suggested in **MUST** specify success and error responses.

---

4. REST Basics - Meta information

MUST contain API meta information [218]

API specifications must contain the following OpenAPI meta information to allow for API management:

- `#/info/title` as (unique) identifying, functional descriptive name of the API
- `#/info/version` to distinguish API specifications versions following semantic rules
- `#/info/description` containing a proper description of the API
- `#/info/contact/{name,url,email}` containing the responsible team

Following OpenAPI extension properties **must** be provided in addition:

- `#/info/x-api-id` unique identifier of the API (see rule 215)
- `#/info/x-audience` intended target audience of the API (see rule 219)

MUST use semantic versioning [116]

OpenAPI allows to specify the API specification version in `#/info/version` . To share a common semantic of version information we expect API designers to comply to Semantic Versioning 2.0 rules `1` to `8` and `11` restricted to the format <MAJOR>.<MINOR>.<PATCH> for versions as follows:

- Increment the **MAJOR** version when you make incompatible API changes after having aligned the changes with consumers,
- Increment the **MINOR** version when you add new functionality in a backwards-compatible

manner, and

- Optionally increment the **PATCH** version when you make backwards-compatible bug fixes or editorial changes not affecting the functionality.

Additional Notes:

- **Pre-release** versions (rule 9) and **build metadata** (rule 10) must not be used in API version information.
- While patch versions are useful for fixing typos etc, API designers are free to decide whether they increment it or not.
- API designers should consider to use API version `0.y.z` (rule 4) for initial API design.

Example:

```
openapi: 3.0.1
info:
  title: Parcel Service API
  description: API for <...>
  version: 1.3.7
  <...>
```

MUST provide API identifiers [215]

Each API specification must be provisioned with a globally unique and immutable API identifier. The API identifier is defined in the `info` -block of the OpenAPI specification and must conform to the following definition:

```
/info/x-api-id:
  type: string
  format: urn
  pattern: ^[a-z0-9][a-z0-9-:.]{6,62}[a-z0-9]$
  description: |
    Mandatory globally unique and immutable API identifier. The API
    id allows to track the evolution and history of an API specification
    as a sequence of versions.
```

API specifications will evolve and any aspect of an OpenAPI specification may change. We require API identifiers because we want to support API clients and providers with API lifecycle management features, like change trackability and history or automated backward compatibility checks. The immutable API identifier allows the identification of all API specification versions of an API evolution. By using API semantic version information or API publishing date as order

criteria you get the **version** or **publication history** as a sequence of API specifications.

Note: While it is nice to use human readable API identifiers based on self-managed URNs, it is recommend to stick to a UUID (freshly generated when first creating the API) to relief API designers from any urge of changing the API identifier while evolving the API. **Do not copy an API unless you immediately change the API identifier in it!**

Example:

```
openapi: 3.0.1
info:
  x-api-id: d0184f38-b98d-11e7-9c56-68f728c1ba70
  title: Parcel Service API
  description: API for <...>
  version: 1.5.8
  <...>
```

MUST provide API audience [219]

Each API must be classified with respect to the intended target **audience** supposed to consume the API, to facilitate differentiated standards on APIs for discoverability, changeability, quality of design and documentation, as well as permission granting. We differentiate the following API audience groups with clear organisational and legal boundaries:

component-internal

This is often referred to as a *team internal API* or a *product internal API*. The API consumers with this audience are restricted to applications of the same **functional component** which typically represents a specific **product** with clear functional scope and ownership. All services of a functional component / product are owned by a specific dedicated owner and engineering team(s). Typical examples of component-internal APIs are APIs being used by internal helper and worker services or that support service operation.

business-unit-internal

The API consumers with this audience are restricted to applications of a specific product portfolio owned by the same business unit.

company-internal

The API consumers with this audience are restricted to applications owned by the business units of the same the company (e.g. Zalando company with Zalando SE, Zalando Payments SE & Co. KG. etc.)

external-partner

The API consumers with this audience are restricted to applications of business partners of the company owning the API and the company itself.

external-public

APIs with this audience can be accessed by anyone with Internet access.

Note: a smaller audience group is intentionally included in the wider group and thus does not need to be declared additionally.

The API audience is provided as API meta information in the `info` -block of the OpenAPI specification and must conform to the following specification:

```
/info/x-audience:
  type: string
  x-extensible-enum:
    - component-internal
    - business-unit-internal
    - company-internal
    - external-partner
    - external-public
  description: |
    Intended target audience of the API. Relevant for standards around
    quality of design and documentation, reviews, discoverability,
    changeability, and permission granting.
```

Note: Exactly **one audience** per API specification is allowed. For this reason a smaller audience group is intentionally included in the wider group and thus does not need to be declared additionally. If parts of your API have a different target audience, we recommend to split API specifications along the target audience — even if this creates redundancies (rationale (internal_link)).

Example:

```
openapi: 3.0.1
info:
  x-audience: company-internal
  title: Parcel Helper Service API
  description: API for <...>
  version: 1.2.4
  <...>
```

For details and more information on audience groups see the API Audience narrative

(internal_link).

MUST/SHOULD/MAY use functional naming schema [223]

Functional naming is a powerful, yet easy way to align global resources as *host*, *permission*, and *event names* within an application landscape. It helps to preserve uniqueness of names while giving readers meaningful context information about the addressed component. Besides, the most important aspect is, that it allows to keep APIs stable in the case of technical and organizational changes (Zalando for example maintains an internal naming convention).

A unique `functional-name` is assigned to each functional component serving an API. It is built of the domain name of the functional group the component is belonging to and a unique a short identifier for the functional component itself:

```
<functional-name>      ::= <functional-domain>-<functional-component>
<functional-domain>    ::= [a-z][a-z0-9-]* -- managed functional group of components
<functional-component> ::= [a-z][a-z0-9-]* -- name of API owning functional component
```

Depending on the API audience, you **must/should/may** follow the functional naming schema for hostnames and event names (and also permission names, in future) as follows:

Functional Naming	Audience
must	external-public, external-partner
should	company-internal, business-unit-internal
may	component-internal

Please see the following rules for detailed functional naming patterns: * **MUST** follow naming convention for hostnames * **MUST** follow naming convention for event type names

Internal Guideance: You *must* use the simple functional name registry (internal_link) to register your functional name before using it. The registry is a centralized infrastructure service to ensure uniqueness of your functional names (and available domains — including subdomains) and to support hostname DNS resolution.
Hint: Due to lexicalic restrictions of DNS names there is no specific separator to split a functional name into (sub) domain and component; this knowledge is only managed in the registry.

MUST follow naming convention for hostnames [224]

Hostnames in APIs must, respectively should conform to the functional naming depending on the audience as follows (see MUST/SHOULD/MAY use functional naming schema for details and `<functional-name>` definition):

```
<hostname>            ::= <functional-hostname> | <application-hostname>

<functional-hostname>  ::= <functional-name>.zalandoapis.com
```

Hint: The following convention (e.g. used by legacy STUPS infrastructure) is deprecated and **only** allowed for hostnames of component-internal APIs:

```
<application-hostname> ::= <application-id>.<organization-unit>.zalan.do
<application-id>       ::= [a-z][a-z0-9-]*  -- application identifier
<organization-id>      ::= [a-z][a-z0-9-]*  -- organization unit identifier, e.g. team identifier
```

Exception: There are legacy hostnames used for APIs with `external-partner` audience which may not follow this rule due to backward compatibility constraints. The API Linter maintains an allow-list for these exceptions (including e.g. `api.merchants.zalando.com` and `api-sandbox.merchants.zalando.com`).

---

5. REST Basics - Security

MUST secure endpoints [104]

Every API endpoint must be protected and armed with authentication and authorization. As part of the API definition you must specify how you protect your API using either the `http` typed `bearer` or `oauth2` typed security schemes defined in the OpenAPI Authentication Specification.

The majority of our APIs (especially the company internal APIs) are protected using JWT tokens provided by the platform IAM token service. In these situations you should use the `http` typed Bearer Authentication security scheme — it is based on OAuth2.0 RFC 6750 defining the standard header `Auhorization: Bearer <token>`. The following code snippet shows how to define the bearer security scheme.

```yaml
components:
  securitySchemes:
    BearerAuth:
```

```
      type: http
      scheme: bearer
      bearerFormat: JWT
```

The bearer security schema can then be applied to all API endpoints, e.g. requiring the token to have `api-repository.read` scope for permission as follows (see also **MUST** define and assign permissions (scopes)):

```
security:
  - BearerAuth: [ api-repository.read ]
```

In other, more specific situations e.g. with customer and partner facing APIs you may use other OAuth 2.0 authorization flows as defined by RFC 6749. Please consult the OpenAPI OAuth 2.0 Authentication section for details on how to define `oauth2` typed security schemes correctly.

Note: Do not use OpenAPI `oauth2` typed security scheme flows (e.g. `implicit`) if your service does not fully support it and implements a simple bearer token scheme, because it exposes authentication server address details and may make use of redirection.

MUST define and assign permissions (scopes) [105]

Endpoints must be equipped with permissions, if they require client authorization for protection since e.g. data is exposed that is classified as `orange` or `red` according to Zalando's Data Classification Group Policy (internal link). Please refer to **MUST** follow the naming convention for permissions (scopes) for designing permission names. Some API endpoints may not require specific permissions for authorization e.g. in case of (i) authorization is *not* needed for the endpoint since all exposed data is classified as `green` or `yellow` , or in case of (ii) the specific authorization is provided differently on the individual object level. In these situations, however, you must make it explicit by assigning the `uid` pseudo permission, which is always available as OAuth2 default scope for all clients in Zalando.

The defined permissions are assigned to each API endpoint based on the security schema (see example in previous section) by specifying the security requirement as follows:

```
paths:
 /business-partners/{partner-id}:
    get:
      summary: Retrieves information about a business partner
      security:
        - BearerAuth: [ business-partner-service.read ]
```

Hint: Following a minimal API specification approach, the `Authorization` -header does not need to be defined on each API endpoint, since it is required and so to say implicitly defined via the security section.

MUST follow the naming convention for permissions (scopes) [225]

As long as the functional naming is not yet supported by our permission registry, permission names in APIs must conform to the following naming pattern:

```
<permission> ::= <standard-permission> |  -- should be sufficient for majority of use cases
                 <resource-permission> |  -- for special security access differentiation use cases
                 <pseudo-permission>      -- used to explicitly indicate that access is not restricted

<standard-permission> ::= <application-id>.<access-mode>
<resource-permission> ::= <application-id>.<resource-name>.<access-mode>
<pseudo-permission>   ::= uid

<application-id>      ::= [a-z][a-z0-9-]*  -- application identifier
<resource-name>       ::= [a-z][a-z0-9-]*  -- free resource identifier
<access-mode>         ::= read | write    -- might be extended in future
```

Note: This naming convention only applies to scopes for service-to-service communication using the Platform IAM tokens. For IAM systems with other naming rules (e.g. Zalando Partner IAM), the naming should be consistent with the existing conventions of those systems.

The permission naming schema corresponds to the naming schema for hostnames and event type names, and typical examples are:

Application ID	Resource ID	Access Type	Example
`order-management`	`sales-order`	`read`	`order-management.sales-order.read`
`order-management`	`shipment-order`	`read`	`order-management.shipment-order.read`
`fulfillment-order`		`write`	`fulfillment-order.write`

Application ID	Resource ID	Access Type	Example
`business-partner-service`		`read`	`business-partner-service.read`

Note: APIs should stick to component specific permissions without resource extension to avoid the complexity of too many fine grained permissions. For the majority of use cases, restricting access for specific API endpoints using read or write is sufficient.

---

6. REST Basics - Data formats

MUST use standard data formats [238]

Open API (based on JSON Schema Validation vocabulary) defines formats from ISO and IETF standards for date/time, integers/numbers and binary data. You **must** use these formats, whenever applicable:

OpenAPI type	OpenAPI format	Specification	Example
`integer`	`int32`	4 byte signed integer between -2^{31} and $2^{31}-1$	`7721071004`
`integer`	`int64`	8 byte signed integer between -2^{63} and $2^{63}-1$	`772107100456824`
`integer`	`bigint`	arbitrarily large signed integer number	`77210710045682438959`
`number`	`float`	`binary32` single precision decimal number — see IEEE 754-2008/ISO 60559:2011	`3.1415927`
`number`	`double`	`binary64` double precision decimal number — see	`3.141592653589793`

OpenAPI type	OpenAPI format	Specification	Example
		IEEE 754-2008/ISO 60559:2011	
number	decimal	arbitrarily precise signed decimal number	3.14159265358979323846264338327
string	byte	base64url encoded byte following RFC 7493 Section 4.4	"VA=="
string	binary	base64url encoded byte sequence following RFC 7493 Section 4.4	"VGVzdA=="
string	date	RFC 3339 internet profile — subset of ISO 8601	"2019-07-30"
string	date-time	RFC 3339 internet profile — subset of ISO 8601	"2019-07-30T06:43:40.252Z"
string	time	RFC 3339 internet profile — subset of ISO 8601	"06:43:40.252Z"
string	duration	RFC 3339 — subset of ISO 8601, see also rule #127 for details.	"P1DT3H4S"
string	period	RFC 3339 — subset of ISO 8601, see also rule #127 for details.	"2022-06-30T14:52:44.276/PT48H" "PT24H/2023-07-30T18:22:16.315Z" "2024-05-15T09:48:56.317Z/.."
string	password		"secret"
string	email	RFC 5322	"example@zalando.de"
string	idn-email	RFC 6531	"hello@bücher.example"

OpenAPI type	OpenAPI format	Specification	Example
string	hostname	RFC 1034	"www.zalando.de"
string	idn-hostname	RFC 5890	"bücher.example"
string	ipv4	RFC 2673	"104.75.173.179"
string	ipv6	RFC 4291	"2600:1401:2::8a"
string	uri	RFC 3986	"https://www.zalando.de/"
string	uri-reference	RFC 3986	"/clothing/"
string	uri-template	RFC 6570	"/users/{id}"
string	iri	RFC 3987	"https://bücher.example/"
string	iri-reference	RFC 3987	"/damenbekleidung-jacken-mäntel/"
string	uuid	RFC 4122	"e2ab873e-b295-11e9-9c02-…"
string	json-pointer	RFC 6901	"/items/0/id"
string	relative-json-pointer	Relative JSON pointers	"1/id"
string	regex	regular expressions as defined in ECMA 262	"^[a-z0-9]+$"

Note: Formats `bigint` and `decimal` have been added to the OpenAPI defined formats — see also **MUST** define a format for number and integer types and **MUST** use standard formats for date and time properties below.

We add further OpenAPI formats that are useful especially in an e-commerce environment e.g. `language code`, `country code`, and `currency` based other ISO and IETF standards. You **must** use these formats, whenever applicable:

OpenAPI type	format	Specification	Example
`string`	`iso-639-1`	two letter language code — see ISO 639-1. Hint: In the past we used `iso-639` as format.	`"en"`
`string`	`bcp47`	multi letter language tag — see BCP 47. It is a compatible extension of ISO 639-1 optionally with additional information for language usage, like region, variant, script.	`"en-DE"`
`string`	`iso-3166-alpha-2`	two letter country code — see ISO 3166-1 alpha-2. Hint: In the past we used `iso-3166` as format.	`"GB"` **Hint:** It is `"GB"`, not `"UK"`, even though `"UK"` has seen some use at Zalando.
`string`	`iso-4217`	three letter currency code — see ISO 4217	`"EUR"`
`string`	`gtin-13`	Global Trade Item Number — see GTIN	`"5710798389878"`

Remark: Please note that this list of standard data formats is not exhaustive and everyone is encouraged to propose additions.

MUST define a format for number and integer types [171]

In **MUST** use standard data formats we added `bigint` and `decimal` to the OpenAPI defined formats. As an implication, you must always provide one of the formats `int32`, `int64`, `bigint` or `float`, `double`, `decimal` when you define an API property of JSON type `number` or `integer`.

By this we prevent clients from guessing the precision incorrectly, and thereby changing the value unintentionally. The precision must be translated by clients and servers into the most specific language types; in Java, for instance, the `number` type with `decimal` format will translate into `BigDecimal` and `integer` type with `int32` format will translate to `int` or `Integer` Java types.

MUST encode binary data in `base64url`

You may expose binary data. You must use a standard media type and data format, if applicable — see Rule 168. If no standard is available, you must define the binary data as `string` typed property with `binary` format using `base64url` encoding — as also described in **MUST** use standard data formats.

MUST use standard formats for date and time properties [169]

As a specific case of **MUST** use standard data formats, you must use the `string` typed formats `date`, `date-time`, `time`, `duration`, or `period` for the definition of date and time properties. The formats are based on the standard RFC 3339 internet profile -- a subset of ISO 8601.

APIs **MUST** use the upper-case `T` as a separator between date and time and upper-case `Z` at the end when generating dates as opposed to lower-case `t`, `z`, space, or any other character. This is stricter than the date/time format as defined in RFC 3339, which leaves it up to the specification.

Exception: For passing date/time information via standard protocol headers, HTTP RFC 7231 requires to follow the date and time specification used by the Internet Message Format RFC 5322.

As defined by the standard, time zone offset may be used, however, we recommend to only use times based on UTC without local offsets. For example `2015-05-28T14:07:17Z` rather than `2015-05-28T14:07:17+00:00`. From experience we have learned that zone offsets are not easy to understand and often not correctly handled. Note also that zone offsets are different from local times which may include daylight saving time. When it comes to storage, all dates should be consistently stored in UTC without a zone offset. Localization should be done locally by the services that provide user interfaces, if required.

Hint: We discourage using numerical timestamps. It typically creates issues with precision, e.g. whether to represent a timestamp as 1460062925, 1460062925000 or 1460062925.000. Date strings, though more verbose and requiring more effort to parse, avoid this ambiguity.

SHOULD select appropriate one of date or date-time format [255]

When choosing between `date` and `datetime` formats you should take into account the following:

- `date` should be used for properties where no exact point in time is required and day time-range is sufficient, for instance, document dates, birthdays, ETAs (estimated time of arrival). Without further context, `date` implies the time period from midnight to midnight in the local time zone. However, the timezone information can be also provided as an additional context information via other fields indicating location.
- `datetime` should be used in all other cases where an exact point in time is required, for instance, datetimes for supplier advice, specific processing events, fast delivery planning dates. As required in **MUST** use standard formats for date and time properties, `datetime` requires the explicit time zone offset to be provided, which avoids misinterpretations and eliminates the need of an additional context to provide.

SHOULD use standard formats for time duration and interval properties [127]

Properties and that are by design durations and time intervals should be represented as strings formatted as defined by ISO 8601 (RFC 3339 Appendix A contains a grammar for `durations` and `periods` - the latter called time intervals in ISO 8601). ISO 8601:1-2019 defines an extension (`..`) to express open ended time intervals that are very convenient in searches and are included in the below ABNF grammar:

```
dur-second        = 1*DIGIT "S"
dur-minute        = 1*DIGIT "M" [dur-second]
dur-hour          = 1*DIGIT "H" [dur-minute]
dur-time          = "T" (dur-hour / dur-minute / dur-second)
dur-day           = 1*DIGIT "D"
dur-week          = 1*DIGIT "W"
dur-month         = 1*DIGIT "M" [dur-day]
dur-year          = 1*DIGIT "Y" [dur-month]
dur-date          = (dur-day / dur-month / dur-year) [dur-time]
duration          = "P" (dur-date / dur-time / dur-week)

period-explicit   = iso-date-time "/" iso-date-time
period-start      = iso-date-time "/" (duration / "..")
period-end        = (duration / "..") "/" iso-date-time
period            = period-explicit / period-start / period-end
```

A time interval query parameter should use `<time-property>_between` instead of the parameter

pair `<time-property>_before` / `<time-property>_after`, while properties providing a time interval should be named `<time-property>_interval`.

MUST use standard formats for country, language and currency properties [170]

As a specific case of **MUST** use standard data formats you must use the following standard formats:

- Country codes: ISO 3166-1-alpha-2 two letter country codes indicated via format `iso-3166-alpha-2` in the OpenAPI specification.
- Language codes: ISO 639-1 two letter language codes indicated via format `iso-639-1` in the OpenAPI specification.
- Language variant tags: BCP 47 multi letter language tag indicated via format `bcp47` in the OpenAPI specification. (It is a compatible extension of ISO 639-1 with additional optional information for language usage, like region, variant, script)
- Currency codes: ISO 4217 three letter currency codes indicated via format `iso-4217` in the OpenAPI specification.

SHOULD use content negotiation, if clients may choose from different resource representations [244]

In some situations the API supports serving different representations of a specific resource (at the same URL), e.g. JSON, PDF, TEXT, or HTML representations for an invoice resource. You should use content negotiation to support clients specifying via the standard HTTP headers `Accept`, `Accept-Language`, `Accept-Encoding` which representation is best suited for their use case, for example, which language of a document, representation / content format, or content encoding. You **SHOULD** use standard media types like `application/json` or `application/pdf` for defining the content format in the `Accept` header.

SHOULD only use UUIDs if necessary [144]

Generating IDs can be a scaling problem in high frequency and near real time use cases. UUIDs solve this problem, as they can be generated without collisions in a distributed, non-coordinated way and without additional server round trips.

However, they also come with some disadvantages:

- pure technical key without meaning; not ready for naming or name scope conventions that might be helpful for pragmatic reasons, e.g. we learned to use names for product attributes, instead of UUIDs
- less usable, because…
  - cannot be memorized and easily communicated by humans
  - harder to use in debugging and logging analysis
  - less convenient for consumer facing usage
- quite long: readable representation requires 36 characters and comes with higher memory and bandwidth consumption
- not ordered along their creation history and no indication of used id volume
- may be in conflict with additional backward compatibility support of legacy ids

UUIDs should be avoided when not needed for large scale id generation. Instead, for instance, server side support with id generation can be preferred (`POST` on id resource, followed by idempotent `PUT` on entity resource). Usage of UUIDs is especially discouraged as primary keys of master and configuration data, like brand-ids or attribute-ids which have low id volume but widespread steering functionality.

Please be aware that sequential, strictly monotonically increasing numeric identifiers may reveal critical, confidential business information, like order volume, to non-privileged clients.

In any case, we should always use string rather than number type for identifiers. This gives us more flexibility to evolve the identifier naming scheme. Accordingly, if used as identifiers, UUIDs should not be qualified using a format property.

Hint: Usually, random UUID is used - see UUID version 4 in RFC 4122. Though UUID version 1 also contains leading timestamps it is not reflected by its lexicographic sorting. This deficit is addressed by ULID (Universally Unique Lexicographically Sortable Identifier). You may favour ULID instead of UUID, for instance, for pagination use cases ordered along creation time.

---

7. REST Basics - URLs

Guidelines for naming and designing resource paths and query parameters.

SHOULD not use /api as base path [135]

In most cases, all resources provided by a service are part of the public API, and therefore should be made available under the root "/" base path.

If the service should also support non-public, internal APIs — for specific operational support functions, for example — we encourage you to maintain two different API specifications and provide API audience. For both APIs, you should not use `/api` as base path.

We see API's base path as a part of deployment variant configuration. Therefore, this information has to be declared in the server object.

MUST pluralize resource names [134]

Usually, a collection of resource instances is provided (at least the API should be ready here). The special case of a *resource singleton* must be modeled as a collection with cardinality 1 including definition of `maxItems` = `minItems` = 1 for the returned `array` structure to make the cardinality constraint explicit.

Exception: the *pseudo identifier* `self` used to specify a resource endpoint where the resource identifier is provided by authorization information (see **MUST** identify resources and sub-resources via path segments).

MUST use URL-friendly resource identifiers [228]

To simplify encoding of resource IDs in URLs they must match the regex `[a-zA-Z0-9:._\-/]*`. Resource IDs only consist of ASCII strings using letters, numbers, underscore, minus, colon, period, and - on rare occasions - slash.

Note: slashes are only allowed to build and signal resource identifiers consisting of compound keys.

Note: to prevent ambiguities of unnormalized paths resource identifiers must never be empty. Consequently, support of empty strings for path parameters is forbidden.

MUST use kebab-case for path segments [129]

Path segments are restricted to ASCII kebab-case strings matching regex `^[a-z][a-z\-0-9]*$`. The first character must be a lower case letter, and subsequent characters can be a letter, or a dash(`-`), or a number.

Example:

```
/shipment-orders/{shipment-order-id}
```

Hint: kebab-case applies to concrete path segments and not necessarily the names of path parameters.

MUST use normalized paths without empty path segments and trailing slashes [136]

You must not specify paths with duplicate or trailing slashes, e.g. `/customers//addresses` or `/customers/` . As a consequence, you must also not specify or use path variables with empty string values.

Note: Non standard paths have no clear semantics. As a result, behavior for non standard paths varies between different HTTP infrastructure components and libraries. This may leads to ambiguous and unexpected results during request handling and monitoring.

We recommend to implement services robust against clients not following this rule. All services **should** normalize request paths before processing by removing duplicate and trailing slashes. Hence, the following requests should refer to the same resource:

```
GET /orders/{order-id}
GET /orders/{order-id}/
GET /orders//{order-id}
```

Note: path normalization is not supported by all framework out-of-the-box. Services are required to support at least the normalized path while rejecting all alternatives paths, if failing to deliver the same resource.

MUST keep URLs verb-free [141]

The API describes resources, so the only place where actions should appear is in the HTTP methods. In URLs, use only nouns. Instead of thinking of actions (verbs), it's often helpful to think about putting a message in a letter box: e.g., instead of having the verb *cancel* in the url, think of sending a message to cancel an order to the *cancellations* letter box on the server side.

MUST avoid actions — think about resources [138]

REST is all about your resources, so consider the domain entities that take part in web service

interaction, and aim to model your API around these using the standard HTTP methods as operation indicators. For instance, if an application has to lock articles explicitly so that only one user may edit them, create an article lock with `PUT` or `POST` instead of using a lock action.

Request:

```
PUT /article-locks/{article-id}
```

The added benefit is that you already have a service for browsing and filtering article locks.

SHOULD define *useful* resources [140]

As a rule of thumb resources should be defined to cover 90% of all its client's use cases. A *useful* resource should contain as much information as necessary, but as little as possible. A great way to support the last 10% is to allow clients to specify their needs for more/less information by supporting filtering and embedding.

MUST use domain-specific resource names [142]

API resources represent elements of the application's domain model. Using domain-specific nomenclature for resource names helps developers to understand the functionality and basic semantics of your resources. It also reduces the need for further documentation outside the API definition. For example, "sales-order-items" is superior to "order-items" in that it clearly indicates which business object it represents. Along these lines, "items" is too general.

SHOULD model complete business processes [139]

An API should contain the complete business processes containing all resources representing the process. This enables clients to understand the business process, foster a consistent design of the business process, allow for synergies from description and implementation perspective, and eliminates implicit invisible dependencies between APIs.

In addition, it prevents services from being designed as thin wrappers around databases, which normally tends to shift business logic to the clients.

MUST identify resources and sub-resources via path segments [143]

Some API resources may contain or reference sub-resources. Embedded sub-resources, which

are not top-level resources, are parts of a higher-level resource and cannot be used outside of its scope. Sub-resources should be referenced by their name and identifier in the path segments as follows:

```
/resources/{resource-id}/sub-resources/{sub-resource-id}
```

In order to improve the consumer experience, you should aim for intuitively understandable URLs, where each sub-path is a valid reference to a resource or a set of resources. For instance, if `/partners/{partner-id}/addresses/{address-id}` is valid, then, in principle, also `/partners/{partner-id}/addresses`, `/partners/{partner-id}` and `/partners` must be valid. Examples of concrete url paths:

```
/shopping-carts/de:1681e6b88ec1/items/1
/shopping-carts/de:1681e6b88ec1
/content/images/9cacb4d8
/content/images
```

Note: resource identifiers may be build of multiple other resource identifiers (see **MAY** expose compound keys as resource identifiers).

Exception: In some situations the resource identifier is not passed as a path segment but via the authorization information, e.g. an authorization token or session cookie. Here, it is reasonable to use `self` as *pseudo-identifier* path segment. For instance, you may define `/employees/self` or `/employees/self/personal-details` as resource paths — and may additionally define endpoints that support identifier passing in the resource path, like define `/employees/{empl-id}` or `/employees/{empl-id}/personal-details`.

MAY expose compound keys as resource identifiers [241]

If a resource is best identified by a *compound key* consisting of multiple other resource identifiers, it is allowed to reuse the compound key in its natural form containing slashes instead of *technical* resource identifier in the resource path without violating the above rule **MUST** identify resources and sub-resources via path segments as follows:

```
/resources/{compound-key-1}[delim-1]...[delim-n-1]{compound-key-n}
```

Example paths:

```
/shopping-carts/{country}/{session-id}
/shopping-carts/{country}/{session-id}/items/{item-id}
```

```
/api-specifications/{docker-image-id}/apis/{path}/{file-name}
/api-specifications/{repository-name}/{artifact-name}:{tag}
/article-size-advices/{sku}/{sales-channel}
```

Note: Exposing a compound key as described above limits ability to evolve the structure of the resource identifier as it is no longer opaque.

To compensate for this drawback, APIs must apply a compound key abstraction consistently in all requests and responses parameters and attributes allowing consumers to treat these as *technical resource identifier* replacement. The use of independent compound key components must be limited to search and creation requests, as follows:

```
# compound key components passed as independent search query parameters
GET /article-size-advices?skus=sku-1,sku-2&sales_channel_id=sid-1
=> { "items": [{ "id": "id-1", ...  },{ "id": "id-2", ...  }] }

# opaque technical resource identifier passed as path parameter
GET /article-size-advices/id-1
=> { "id": "id-1", "sku": "sku-1", "sales_channel_id": "sid-1", "size": ... }

# compound key components passed as mandatory request fields
POST /article-size-advices { "sku": "sku-1", "sales_channel_id": "sid-1", "size": ... }
=> { "id": "id-1", "sku": "sku-1", "sales_channel_id": "sid-1", "size": ... }
```

Where **`id-1`** is representing the opaque provision of the compound key **`sku-1/sid-1`** as technical resource identifier.

Remark: A compound key component may itself be used as another resource identifier providing another resource endpoint, e.g **`/article-size-advices/{sku}`** .

MAY consider using (non-) nested URLs [145]

If a sub-resource is only accessible via its parent resource and may not exist without parent resource, consider using a nested URL structure, for instance:

```
/shoping-carts/de/1681e6b88ec1/cart-items/1
```

However, if the resource can be accessed directly via its unique id, then the API should expose it as a top-level resource. For example, customer has a collection for sales orders; however, sales orders have globally unique id and some services may choose to access the orders directly, for instance:

```
/customers/1637asikzec1
/sales-orders/5273gh3k525a
```

SHOULD limit number of resource types [146]

To keep maintenance and service evolution manageable, we should follow "functional segmentation" and "separation of concern" design principles and do not mix different business functionalities in same API definition. In practice this means that the number of resource types exposed via an API should be limited. In this context a resource type is defined as a set of highly related resources such as a collection, its members and any direct sub-resources.

For example, the resources below would be counted as three resource types, one for customers, one for the addresses, and one for the customers' related addresses:

```
/customers
/customers/{id}
/customers/{id}/preferences
/customers/{id}/addresses
/customers/{id}/addresses/{addr}
/addresses
/addresses/{addr}
```

Note that:

- We consider `/customers/id/preferences` part of the `/customers` resource type because it has a one-to-one relation to the customer without an additional identifier.
- We consider `/customers` and `/customers/id/addresses` as separate resource types because `/customers/id/addresses/{addr}` also exists with an additional identifier for the address.
- We consider `/addresses` and `/customers/id/addresses` as separate resource types because there’s no reliable way to be sure they are the same.

Given this definition, our experience is that well defined APIs involve no more than 4 to 8 resource types. There may be exceptions with more complex business domains that require more resources, but you should first check if you can split them into separate subdomains with distinct APIs.

Nevertheless one API should hold all necessary resources to model complete business processes helping clients to understand these flows.

SHOULD limit number of sub-resource levels [147]

There are main resources (with root url paths) and sub-resources (or *nested* resources with non-root urls paths). Use sub-resources if their life cycle is (loosely) coupled to the main resource, i.e. the main resource works as collection resource of the subresource entities. You should use ⇐ 3 sub-resource (nesting) levels — more levels increase API complexity and url path length. (Remember, some popular web browsers do not support URLs of more than 2000 characters.)

MUST use snake_case (never camelCase) for query parameters [130]

See also **MUST** property names must be snake_case (and never camelCase).

MUST stick to conventional query parameters [137]

If you provide query support for searching, sorting, filtering, and paginating, you must stick to the following naming conventions:

- `q` : default query parameter, e.g. used by browser tab completion; should have an entity specific alias, e.g. sku.
- `sort` : comma-separated list of fields (as defined by **MUST** define collection format of header and query parameters) to define the sort order. To indicate sorting direction, fields may be prefixed with `+` (ascending) or `-` (descending), e.g. /sales-orders?sort=+id.
- `fields` : field name expression to retrieve only a subset of fields of a resource. See **SHOULD** support partial responses via filtering below.
- `embed` : field name expression to expand or embedded sub-entities, e.g. inside of an article entity, expand silhouette code into the silhouette object. Implementing `embed` correctly is difficult, so do it with care. See **SHOULD** allow optional embedding of sub-resources below.
- `offset` : numeric offset of the first element provided on a page representing a collection request. See REST Design - Pagination section below.
- `cursor` : an opaque pointer to a page, never to be inspected or constructed by clients. It usually (encrypted) encodes the page position, i.e. the identifier of the first or last page element, the pagination direction, and the applied query filters to recreate the collection. See Cursor-based pagination in RESTful APIs or REST Design - Pagination section below.
- `limit` : client suggested limit to restrict the number of entries on a page. See REST Design - Pagination section below.

8. REST Basics - JSON payload

These guidelines provides recommendations for defining JSON data at Zalando. JSON here refers to RFC 7159 (which updates RFC 4627), the "application/json" media type and custom JSON media types defined for APIs. The guidelines clarifies some specific cases to allow Zalando JSON data to have an idiomatic form across teams and services.

MUST use JSON as payload data interchange format [167]

Use JSON (RFC 7159) to represent structured (resource) data passed with HTTP requests and responses as body payload. The JSON payload must use a JSON object as top-level data structure (if possible) to allow for future extension. This also applies to collection resources, where you ad-hoc would use an array — see also **MUST** always return JSON objects as top-level data structures.

Additionally, the JSON payload must comply to the more restrictive Internet JSON (RFC 7493), particularly

- Section 2.1 on encoding of characters, and
- Section 2.3 on object constraints.

As a consequence, a JSON payload must

- use `UTF-8` encoding
- consist of valid Unicode strings, i.e. must not contain non-characters or surrogates, and
- contain only unique member names (no duplicate names).

SHOULD design single resource schema for reading and writing [252]

To simplify API specs and API maintenance, as well as the cognitive load of developers, endpoints should utilize a common model for reading and writing the same resource type. The differences between read and write models are best addressed by

- marking properties `writeOnly` that are only available in a request, e.g. passwords, and
- marking properties `readOnly` that are only available in a response, e.g. resource identifiers.

Note: A resource with property marked `readOnly` in the API according to the JSON schema validation may be ignored or rejected by a server. Both are in line with the idea behind the compatibility guidance, however, we suggest to document when rejecting to prevent client side surprises.

SHOULD be aware of services not fully supporting JSON/unicode [250]

JSON payloads passed via API requests consist of valid unicode characters (see **MUST** use JSON as payload data interchange format). While `\uxxxx` are valid characters in a JSON strings, they can create failures when leaving the API request processing context, e.g. by writing the data into a database or piping it through command line tools that are not 100% compatible with the JSON or unicode standard.

A prominent example for such a tool is the Postgres database:

- Postgres cannot handle `\u0000` in `jsonb` and `text` types (see datatype-json).

As a consequence, servers and clients that forward JSON content to other tools should check that these tools fully support the JSON or unicode standard. If not, they should reject or sanitize unicode characters not supported by the tools.

Note: A sanitization and rejection are actions that change the API behavior and therefore should be described by the API specification.

MAY pass non-JSON media types using data specific standard formats [168]

Non-JSON media types may be supported, if you stick to a business object specific standard format for the payload data, for instance, image data format (JPG, PNG, GIF), document format (PDF, DOC, ODF, PPT), or archive format (TAR, ZIP).

Generic structured data interchange formats other than JSON (e.g. XML, CSV) may be provided, but only additionally to JSON as default format using content negotiation, for specific use cases where clients may not interpret the payload structure.

SHOULD use standard media types [172]

You should use standard media types (defined in media type registry of Internet Assigned

Numbers Authority (IANA)) as `content-type` (or `accept`) header information. More specifically, for JSON payload you should use the standard media type `application/json` (or `application/problem+json` for **MUST** support problem JSON).

You should avoid using custom media types like `application/x.zalando.article+json`. Custom media types beginning with `x` bring no advantage compared to the standard media type for JSON, and make automated processing more difficult.

Exception: Custom media type should be only used in situations where you need to provide API endpoint versioning (with content negotiation) due to incompatible changes.

SHOULD pluralize array names [120]

Names of arrays should be pluralized to indicate that they contain multiple values. This implies in turn that object names should be singular.

MUST property names must be snake_case (and never camelCase) [118]

Property names are restricted to ASCII snake_case strings matching regex `^[a-z_][a-z_0-9]*$`. The first character must be a lower case letter, or an underscore, and subsequent characters can be a letter, an underscore, or a number.

Examples:

```
customer_number, sales_order_number, billing_address
```

Rationale: No established industry standard exists, but many popular Internet companies prefer snake_case: e.g. GitHub, Stack Exchange, Twitter. Others, like Google and Amazon, use both - but not only camelCase. It's essential to establish a consistent look and feel such that JSON looks as if it came from the same hand.

SHOULD declare enum values using UPPER_SNAKE_CASE string [240]

Enumerations should be represented as `string` typed OpenAPI definitions of request parameters or model properties. Enum values (using `enum` or `x-extensible-enum`) need to consistently use the upper-snake case format, e.g. `VALUE` or `YET_ANOTHER_VALUE`. This approach allows to clearly distinguish values from properties or other elements.

Exception: This rule does not apply for case sensitive values sourced from outside API definition scope, e.g. for language codes from ISO 639-1, or when declaring possible values for a rule 137 [`sort` parameter].

SHOULD use naming convention for date/time properties [235]

Naming of date and data-time properties should comply with the following convention: The name either (i) contains `date`, `day`, `time`, `timestamp` or similar type indicators, or (ii) end with the `_at` suffix. The convention allows easy identification of date/time typed properties and distinguishing them e.g. from boolean properties like `created` vs. `created_at`. Examples of valid date/time property names:

- `created_at`, `modified_at`, `occurred_at`, `returned_at` — instead of `created`, `modified`, …
- `campaign_start_time`, `arrival_date`, `checkout_time` — instead of `campaign_start`, `arrival`, …

Hint: Use `format: date-time` or `format: date` as required in **MUST** use standard data formats.

SHOULD define maps using `additionalProperties` [216]

A "map" here is a mapping from string keys to some other type. In JSON this is represented as an object, the key-value pairs being represented by property names and property values. In OpenAPI schema (as well as in JSON schema) they should be represented using additionalProperties with a schema defining the value type. Such an object should normally have no other defined properties.

The map keys don't count as property names in the sense of rule 118, and can follow whatever format is natural for their domain. Please document this in the description of the map object's schema.

Here is an example for such a map definition (the `translations` property):

```
components:
  schemas:
    Message:
      description:
        A message together with translations in several languages.
      type: object
```

```
    properties:
      message_key:
        type: string
        description: The message key.
      translations:
        description:
          The translations of this message into several languages.
          The keys are [IETF BCP-47 language tags](https://tools.ietf.org/html/bcp47).
        type: object
        additionalProperties:
          type: string
          description:
            the translation of this message into the language identified by the key.
```

An actual JSON object described by this might then look like this:

```
{ "message_key": "color",
  "translations": {
    "de": "Farbe",
    "en-US": "color",
    "en-GB": "colour",
    "eo": "koloro",
    "nl": "kleur"
  }
}
```

MUST use same semantics for `null` and absent properties [123]

OpenAPI 3.x allows to mark properties as **`required`** and as **`nullable`** to specify whether properties may be absent (**`{}`**) or **`null`** (**`{"example":null}`**). If a property is defined to be not **`required`** and **`nullable`** (see 2nd row in Table below), this rule demands that both cases must be handled in the exact same manner by specification.

The following table shows all combinations and whether the examples are valid:

required	**nullable**	**{}**	**{"example":null}**
true	**true**	❌ No	✓ Yes
false	**true**	✓ Yes	✓ Yes
true	**false**	❌ No	❌ No
false	**false**	✓ Yes	❌ No

While API designers and implementers may be tempted to assign different semantics to both cases, we explicitly decide **against** that option, because we think that any gain in expressiveness is far outweighed by the risk of clients not understanding and implementing the subtle differences incorrectly.

As an example, an API that provides the ability for different users to coordinate on a time schedule, e.g. a meeting, may have a resource for options in which every user has to make a `choice`. The difference between *undecided* and *decided against any of the options* could be modeled as *absent* and `null` respectively. It would be safer to express the `null` case with a dedicated Null object, e.g. `{}` compared to `{"id":"42"}`.

Moreover, many major libraries have somewhere between little to no support for a `null`/absent pattern (see Gson, Moshi, Jackson, JSON-B). Especially strongly-typed languages suffer from this since a new composite type is required to express the third state. Nullable `Option`/`Optional`/`Maybe` types could be used but having nullable references of these types completely contradicts their purpose.

The only exception to this rule is JSON Merge Patch RFC 7396) which uses `null` to explicitly indicate property deletion while absent properties are ignored, i.e. not modified.

MUST not use `null` for boolean properties [122]

Schema based JSON properties that are by design booleans must not be presented as nulls. A boolean is essentially a closed enumeration of two values, true and false. If the content has a meaningful null value, we strongly prefer to replace the boolean with enumeration of named values or statuses - for example accepted_terms_and_conditions with enumeration values YES, NO, UNDEFINED.

SHOULD not use `null` for empty arrays [124]

Empty array values can unambiguously be represented as the empty list, `[]`.

MUST use common field names and semantics [174]

You must use common field names and semantics whenever applicable. Common fields are idiomatic, create consistency across APIs and support common understanding for API consumers.

We define the following common field names:

- `id` : the identity of the object. If used, IDs must be opaque strings and not numbers. IDs are unique within some documented context, are stable and don't change for a given object once assigned, and are never recycled cross entities.
- `xyz_id` : an attribute within one object holding the identifier of another object must use a name that corresponds to the type of the referenced object or the relationship to the referenced object followed by `_id` (e.g. `partner_id` not `partner_number` , or `parent_node_id` for the reference to a parent node from a child node, even if both have the type `Node`). **Exception**: We use `customer_number` instead of `customer_id` for customer facing identification of customers due to legacy reasons. (Hint: `customer_id` used to be defined as internal only, technical integer key, see Naming Decision: `customer_number` vs `customer_id` [internal_link]).
- `etag` : the `etag` of an embedded sub-resource. It typically is used to carry the `ETag` for subsequent `PUT` / `PATCH` calls (see `ETag` and `ETags` in result entities).

Further common fields are defined in **SHOULD** use naming convention for date/time properties. The following guidelines define standard objects and fields:

- **SHOULD** use pagination response page object
- **MUST** use the common address fields
- **MUST** use the common money object

Example JSON schema:

```
tree_node:
  type: object
  properties:
    id:
      description: the identifier of this node
      type: string
    parent_node_id:
      description: the identifier of the parent node of this node
      type: string
    created_at:
      description: when got this node created
      type: string
      format: 'date-time'
    modified_at:
      description: when got this node last updated
      type: string
      format: 'date-time'
  example:
    id: '123435'
```

```
    parent_node_id: '534321'
    created_at: '2017-04-12T23:20:50.52Z'
    modified_at: '2017-04-12T23:20:50.52Z'
```

MUST use the common address fields [249]

Address structures play a role in different business and use-case contexts, including country variances. All attributes that relate to address information must follow the naming and semantics defined below.

```
addressee:
  description: a (natural or legal) person that gets addressed
  type: object
  properties:
    salutation:
      description: |
        a salutation and/or title used for personal contacts to some
        addressee; not to be confused with the gender information!
      type: string
      example: Mr
    first_name:
      description: |
        given name(s) or first name(s) of a person; may also include the
        middle names.
      type: string
      example: Hans Dieter
    last_name:
      description: |
        family name(s) or surname(s) of a person
      type: string
      example: Mustermann
    business_name:
      description: |
        company name of the business organization. Used when a business is
        the actual addressee; for personal shipments to office addresses, use
        `care_of` instead.
      type: string
      example: Consulting Services GmbH
  required:
    - first_name
    - last_name

address:
  description:
    an address of a location/destination
  type: object
  properties:
    care_of:
      description: |
        (aka c/o) the person that resides at the address, if different from
```

```
      addressee. E.g. used when sending a personal parcel to the
      office /someone else's home where the addressee resides temporarily
    type: string
    example: Consulting Services GmbH
  street:
    description: |
      the full street address including house number and street name
    type: string
    example: Schönhauser Allee 103
  additional:
    description: |
      further details like building name, suite, apartment number, etc.
    type: string
    example: 2. Hinterhof rechts
  city:
    description: |
      name of the city / locality
    type: string
    example: Berlin
  zip:
    description: |
      zip code or postal code
    type: string
    example: 14265
  country_code:
    description: |
      the country code according to
      [iso-3166-1-alpha-2](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2)
    type: string
    example: DE
required:
  - street
  - city
  - zip
  - country_code
```

Grouping and cardinality of fields in specific data types may vary based on the specific use case (e.g. combining addressee and address fields into a single type when modeling an address label vs distinct addressee and address types when modeling users and their addresses).

MUST use the common money object [173]

Use the following common money structure:

```
Money:
  type: object
  properties:
    amount:
      type: number
```

```
      description: >
        The amount describes unit and subunit of the currency in a single value,
        where the integer part (digits before the decimal point) is for the
        major unit and fractional part (digits after the decimal point) is for
        the minor unit.
      format: decimal
      example: 99.95
    currency:
      type: string
      description: 3 letter currency code as defined by ISO-4217
      format: iso-4217
      example: EUR
  required:
    - amount
    - currency
```

APIs are encouraged to include a reference to the global schema for Money.

```
SalesOrder:
  properties:
    grand_total:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/money-1.0.0.yaml#/Money'
```

Please note that APIs have to treat Money as a closed data type, i.e. it's not meant to be used in an inheritance hierarchy. That means the following usage is not allowed:

```
{
  "amount": 19.99,
  "currency": "EUR",
  "discounted_amount": 9.99
}
```

Cons

- Violates the Liskov Substitution Principle
- Breaks existing library support, e.g. Jackson Datatype Money
- Less flexible since both amounts are coupled together, e.g. mixed currencies are impossible

A better approach is to favor composition over inheritance:

```
{
    "price": {
      "amount": 19.99,
      "currency": "EUR"
    },
```

```
    "discounted_price": {
      "amount": 9.99,
      "currency": "EUR"
    }
}
```

Pros

- No inheritance, hence no issue with the substitution principle
- Makes use of existing library support
- No coupling, i.e. mixed currencies is an option
- Prices are now self-describing, atomic values

Notes

Please be aware that some business cases (e.g. transactions in Bitcoin) call for a higher precision, so applications must be prepared to accept values with unlimited precision, unless explicitly stated otherwise in the API specification.

Examples for correct representations (in EUR):

- **`42.20`** or **`42.2`** = 42 Euros, 20 Cent
- **`0.23`** = 23 Cent
- **`42.0`** or **`42`** = 42 Euros
- **`1024.42`** = 1024 Euros, 42 Cent
- **`1024.4225`** = 1024 Euros, 42.25 Cent

Make sure that you don't convert the "amount" field to **`float`** / **`double`** types when implementing this interface in a specific language or when doing calculations. Otherwise, you might lose precision. Instead, use exact formats like Java's **`BigDecimal`** . See Stack Overflow for more info.

Some JSON parsers (NodeJS's, for example) convert numbers to floats by default. After discussing the pros and cons we've decided on "decimal" as our amount format. It is not a standard OpenAPI format, but should help us to avoid parsing numbers as float / doubles.

9. REST Basics - HTTP requests

MUST use HTTP methods correctly [148]

Be compliant with the standardized HTTP semantics (see RFC-9110 "HTTP semantics") as summarized here.

GET

`GET` requests are used to **read** either a single or a collection resource.

- `GET` requests for individual resources will usually generate a **404** (if the resource does not exist).
- `GET` requests for collection resources may return either **200** (if the collection is empty) or **404** (if the collection is missing).
- `GET` requests must NOT have a request body payload (see `GET with body`).

Note: `GET` requests on collection resources should provide sufficient filter and REST Design - Pagination mechanisms.

GET with body payload

APIs sometimes face the problem, that they have to provide extensive structured request information with `GET`, that may conflict with the size limits of clients, load-balancers, and servers. As we require APIs to be standard conform (request body payload in `GET` must be ignored on server side), API designers have to check the following two options:

1. `GET` with URL encoded query parameters: when it is possible to encode the request information in query parameters, respecting the usual size limits of clients, gateways, and servers, this should be the first choice. The request information can either be provided via multiple query parameters or by a single structured URL encoded string.
2. `POST` with body payload content: when a `GET` with URL encoded query parameters is not possible, a `POST` request with body payload must be used, and explicitly documented with a hint like in the following example:

```
paths:
  /products:
    post:
```

```
    description: >
      [GET with body payload](https://opensource.zalando.com/restful-api-guidelines/#get-with-body)
      - no resources created: Returns all products matching the query passed
      as request input payload.
    requestBody:
      required: true
      content:
        ...
```

Note: It is no option to encode the lengthy structured request information using header parameters. From a conceptual point of view, the semantic of an operation should always be expressed by the resource names, as well as the involved path and query parameters. In other words by everything that goes into the URL. Request headers are reserved for general context information (see **SHOULD** use only the specified proprietary Zalando headers). In addition, size limits on query parameters and headers are not reliable and depend on clients, gateways, server, and actual settings. Thus, switching to headers does not solve the original problem.

Hint: As `GET with body` is used to transport extensive query parameters, the `cursor` cannot any longer be used to encode the query filters in case of cursor-based pagination. As a consequence, it is best practice to transport the query filters in the body payload, while keeping the pagination parameters (`cursor` , `limit`) in query parameters. This allows using pagination links containing the `cursor` that is only encoding the page position and direction. To protect the pagination sequence, the `cursor` may also contain a hash over all applied query filters, which is then validated against the request body. (See also **SHOULD** use pagination links.)

PUT

`PUT` requests are used to **update** (and sometimes to create) **entire** resources – single or collection resources. The semantic is best described as *"please put the enclosed representation at the resource mentioned by the URL, replacing any existing resource."*.

- `PUT` requests are usually applied to single resources, and not to collection resources, as this would imply replacing the entire collection.
- `PUT` requests are usually robust against non-existence of resources by implicitly creating the resource before updating.
- On successful `PUT` requests, the server will **replace the entire resource** addressed by the URL with the representation passed in the payload. Subsequent reads will deliver the same payload, plus possibly server-generated fields like `modified_at` .
- Successful `PUT` requests return **200** or **204** (if the resource was updated — with or without

returning the resource), **201** (if the resource was newly created), and **202** (if the request was accepted for asynchronous processing).

The updated/created resource may be returned as response payload. We recommend, to not use it per default, but if the resource is enriched with server-generated fields like `version_number`. You may also support client side steering (see **MAY** consider to support `Prefer` header to handle processing preferences).

Important: It is good practice to keep the resource identifier management under control of the service provider and not the client, and, hence, to prefer `POST` for creation of (at least top-level) resources, and focus `PUT` on its usage for updates. However, in situations where the identifier and all resource attributes are under control of the client as input for the resource creation you should use `PUT` and pass the resource identifier as URL path parameter. Putting the same resource twice is required to be idempotent and to result in the same single resource instance (see **MUST** fulfill common method properties) without data duplication in case of repetition.

Hint: To prevent unnoticed concurrent updates and duplicate creations when using `PUT`, you **MAY** consider to support `ETag` together with `If-Match` / `If-None-Match` header to allow the server to react on stricter demands that expose conflicts and prevent lost updates. See also Optimistic locking in RESTful APIs for details and options.

POST

`POST` requests are idiomatically used to **create** single resources on a collection resource endpoint, but other semantics on single resources endpoint are equally possible. The semantic for collection endpoints is best described as *"please add the enclosed representation to the collection resource identified by the URL"*. The semantic for single resource endpoints is best described as *"please execute the given well specified request on the resource identified by the URL"*.

- Successful `POST` requests on a collection endpoint will create one or more resource instances.
- For single resources, you **SHOULD** return **201** and the new resource object (including the resource identifier) in the response payload.
  - The client **SHOULD** be able to use this resource identifier to `GET` the resource if supported by the API (see **MUST** identify resources and sub-resources via path segments).
  - The URL to `GET` the resource **SHOULD** be provided in the `Location` header.

  - If the response does not contain the created resource or a resource monitor, the status code **SHOULD** be **201** and the URL to `GET` the resource **MUST** be provided in the `Location` response header.
  - If the `POST` is idempotent, the status code **SHOULD** be **201** if the resource was created and **200** or **204** if the resource was updated.

The resource identifier **MUST NOT** be passed as request body data by the client, but created and maintained by the service.

- For multiple resources you **SHOULD** return **201** in the response, as long as they are created atomically, meaning that either all the resources are created or none of them are.
  - You **MUST** use code 207 for batch or bulk requests if the request can partially fail, that is some of the resources can fail to be created.

Posting the same resource twice is **not** required to be idempotent (check **MUST** fulfill common method properties) and may result in multiple resources. However, you **SHOULD** consider to design `POST` and `PATCH` idempotent to prevent this.

- You **MAY** support asynchronous request processing, where the resource creation would not finish by the time the request is delivered, so the response status code **SHOULD** be **202**.

Apart from resource creation, `POST` should be also used for scenarios that cannot be covered by the other methods sufficiently. However, in such cases make sure to document the fact that `POST` is used as a workaround (see e.g. `GET with body`).

PATCH

`PATCH` method extends HTTP via RFC-5789 standard to update parts of the resource objects where e.g. in contrast to `PUT` only a specific subset of resource fields should be changed. The set of changes is represented in a format called a *patch document* passed as payload and identified by a specific media type. The semantic is best described as *"please change the resource identified by the URL according to my patch document"*. The syntax and semantics of the patch document is not defined in RFC-5789 and must be described in the API specification by using specific media types.

- `PATCH` requests are usually applied to single resources as patching entire collection is challenging.

- `PATCH` requests are usually not robust against non-existence of resource instances.
- On successful `PATCH` requests, the server will update parts of the resource addressed by the URL as defined by the change request in the payload.
- Successful `PATCH` requests return **200** or **204** (if the resource was updated — with or without returning the resource), and **202** (if the request was accepted for asynchronous processing).

Note: since implementing `PATCH` correctly is a bit tricky, we strongly suggest to choose one and only one of the following patterns per endpoint (unless forced by a backwards compatible change). In preference order:

1. Use `PUT` with complete objects to update a resource as long as feasible (i.e. do not use `PATCH` at all). **Note:** this choice by the API server imposes additional requirements on the client (**MUST** prepare clients to accept compatible API extensions) which can be implemented e.g. in Java following the practice Handling compatible API extensions.
2. Use `PATCH` with JSON Merge Patch standard, a specialized media type `application/merge-patch+json` for partial resource representation to update parts of resource objects.
3. Use `PATCH` with JSON Patch standard, a specialized media type `application/json-patch+json` that includes instructions on how to change the resource.
4. Use `POST` (with a proper description of what is happening) instead of `PATCH`, if the request does not modify the resource in a way defined by the semantics of the standard media types above.

In practice JSON Merge Patch quickly turns out to be too limited, especially when trying to update single objects in large collections (as part of the resource). In this case JSON Patch is more powerful while still showing readable patch requests (see also JSON patch vs. merge). JSON Patch supports changing of array elements identified via its index, but not via (key) fields of the elements as typically needed for collections.

Note: Patching the same resource twice is **not** required to be idempotent (check **MUST** fulfill common method properties) and may result in a changing result. However, you **SHOULD** consider to design `POST` and `PATCH` idempotent to prevent this.

Hint: To prevent unnoticed concurrent updates when using `PATCH` you **MAY** consider to support `ETag` together with `If-Match` / `If-None-Match` header to allow the server to react on stricter demands that expose conflicts and prevent lost updates. See Optimistic locking in RESTful APIs and **SHOULD** consider to design `POST` and `PATCH` idempotent for details and options.

DELETE

`DELETE` requests are used to **delete** resources. The semantic is best described as *"please delete the resource identified by the URL"*.

- `DELETE` requests are usually applied to single resources, not on collection resources, as this would imply deleting the entire collection.
- `DELETE` request can be applied to multiple resources at once using query parameters on the collection resource (see DELETE with query parameters).
- Successful `DELETE` requests return **200** or **204** (if the resource was deleted — with or without returning the resource), or **202** (if the request was accepted for asynchronous processing).
- Failed `DELETE` requests will usually generate **404** (if the resource cannot be found) or **410** (if the resource was already traceably deleted before).

Important: After deleting a resource with `DELETE`, a `GET` request on the resource is expected to either return **404** (not found) or **410** (gone) depending on how the resource is represented after deletion. Under no circumstances the resource must be accessible after this operation on its endpoint.

DELETE with query parameters

`DELETE` request can have query parameters. Query parameters should be used as filter parameters on a resource and not for passing context information to control the operation behavior.

```
DELETE /resources?param1=value1&param2=value2...&paramN=valueN
```

Note: When providing `DELETE` with query parameters, API designers must carefully document the behavior in case of (partial) failures to manage client expectations properly.

The response status code of `DELETE` with query parameters requests should be similar to usual `DELETE` requests. In addition, it may return the status code **207** using a payload describing the operation results (see **MUST** use code 207 for batch or bulk requests for details).

DELETE with body payload

In rare cases `DELETE` may require additional information, that cannot be classified as filter parameters and thus should be transported via request body payload, to perform the operation. Since RFC-9110 Section 9.3.5 states, that `DELETE` has an undefined semantic for payloads, we

recommend to utilize `POST`. In this case the POST endpoint must be documented with the hint `DELETE with body` analog to how it is defined for `GET with body`. The response status code of `DELETE with body` requests should be similar to usual `DELETE` requests.

HEAD

`HEAD` requests are used to **retrieve** the header information of single resources and resource collections.

- `HEAD` has exactly the same semantics as `GET`, but returns headers only, no body.

Hint: `HEAD` is particular useful to efficiently lookup whether large resources or collection resources have been updated in conjunction with the `ETag`-header.

OPTIONS

`OPTIONS` requests are used to **inspect** the available operations (HTTP methods) of a given endpoint.

- `OPTIONS` responses usually either return a comma separated list of methods in the `Allow` header or as a structured list of link templates.

Note: `OPTIONS` is rarely implemented, though it could be used to self-describe the full functionality of a resource.

MUST fulfill common method properties [149]

Request methods in RESTful services can be…

- safe — the operation semantic is defined to be read-only, meaning it must not have *intended side effects*, i.e. changes, to the server state.
- idempotent — the operation has the same *intended effect* on the server state, independently whether it is executed once or multiple times. **Note:** this does not require that the operation is returning the same response or status code.
- cacheable — to indicate that responses are allowed to be stored for future reuse. In general, requests to safe methods are cacheable, if it does not require a current or authoritative response from the server.

Note: The above definitions, of *intended (side) effect* allows the server to provide additional state changing behavior as logging, accounting, pre- fetching, etc. However, these actual effects and

state changes, must not be intended by the operation so that it can be held accountable.

Method implementations must fulfill the following basic properties according to RFC 9110 Section 9.2:

Method	Safe	Idempotent	Cacheable
`GET`	✓ Yes	✓ Yes	✓ Yes
`HEAD`	✓ Yes	✓ Yes	✓ Yes
`POST`	✗ No	⚠ No, but **SHOULD** consider to design `POST` and `PATCH` idempotent	⚠ May, but only if specific `POST` endpoint is safe. **Hint:** not supported by most caches.
`PUT`	✗ No	✓ Yes	✗ No
`PATCH`	✗ No	⚠ No, but **SHOULD** consider to design `POST` and `PATCH` idempotent	✗ No
`DELETE`	✗ No	✓ Yes	✗ No
`OPTIONS`	✓ Yes	✓ Yes	✗ No
`TRACE`	✓ Yes	✓ Yes	✗ No

Note: MUST document cacheable `GET`, `HEAD`, and `POST` endpoints.

SHOULD consider to design `POST` and `PATCH` idempotent [229]

In many cases it is helpful or even necessary to design `POST` and `PATCH` idempotent for clients to expose conflicts and prevent resource duplicate (a.k.a. zombie resources) or lost updates, e.g. if same resources may be created or changed in parallel or multiple times. To design an idempotent API endpoint owners should consider to apply one of the following three patterns.

- A resource specific **conditional key** provided via `If-Match` header in the request. The key is in general a meta information of the resource, e.g. a *hash* or *version number*, often stored with it. It allows to detect concurrent creations and updates to ensure idempotent behavior (see **MAY** consider to support `ETag` together with `If-Match` / `If-None-Match` header).

- A resource specific **secondary key** provided as resource property in the request body. The *secondary key* is stored permanently in the resource. It allows to ensure idempotent behavior by looking up the unique secondary key in case of multiple independent resource creations from different clients (see **SHOULD** use secondary key for idempotent `POST` design).
- A client specific **idempotency key** provided via `Idempotency-Key` header in the request. The key is not part of the resource but stored temporarily pointing to the original response to ensure idempotent behavior when retrying a request (see **MAY** consider to support `Idempotency-Key` header).

Note: While **conditional key** and **secondary key** are focused on handling concurrent requests, the **idempotency key** is focused on providing the exact same responses, which is even a *stronger* requirement than the idempotency defined above. It can be combined with the two other patterns.

To decide, which pattern is suitable for your use case, please consult the following table showing the major properties of each pattern:

	Conditional Key	**Secondary Key**	**Idempotency Key**
Applicable with	`PATCH`	`POST`	`POST` / `PATCH`
HTTP Standard	✓ Yes	✗ No	✗ No
Prevents duplicate (zombie) resources	✓ Yes	✓ Yes	✗ No
Prevents concurrent lost updates	✓ Yes	✗ No	✗ No
Supports safe retries	✓ Yes	✓ Yes	✓ Yes
Supports exact same response	✗ No	✗ No	✓ Yes
Can be inspected (by intermediaries)	✓ Yes	✗ No	✓ Yes
Usable without previous `GET`	✗ No	✓ Yes	✓ Yes

Note: The patterns applicable to `PATCH` can be applied in the same way to `PUT` and `DELETE` providing the same properties.

If you mainly aim to support safe retries, we suggest to apply conditional key and secondary key pattern before the idempotency key pattern.

Note: like for `PUT`, successful `POST` or `PATCH` returns **200** or **204** (if the resource was updated — with or without returning the resource), or **201** (if resource was created). Hence, clients can differentiate successful robust repetition from resource created server activity of idempotent `POST`.

SHOULD use secondary key for idempotent `POST` design [231]

The most important pattern to design `POST` idempotent for creation is to introduce a resource specific **secondary key** provided in the request body, to eliminate the problem of duplicate (a.k.a zombie) resources.

The secondary key is stored permanently in the resource as *alternate key* or *combined key* (if consisting of multiple properties) guarded by a uniqueness constraint enforced server-side, that is visible when reading the resource. The best and often naturally existing candidate is a *unique foreign key*, that points to another resource having *one-on-one* relationship with the newly created resource, e.g. a parent process identifier.

A good example here for a secondary key is the shopping cart ID in an order resource.

Note: When using the secondary key pattern without `Idempotency-Key` all subsequent retries should fail with status code **409** (conflict). We suggest to avoid **200** here unless you make sure, that the delivered resource is the original one implementing a well defined behavior. Using **204** without content would be a similar well defined option.

MAY support asynchronous request processing [253]

Typically REST APIs are designed as synchronous interfaces where all server-side processing and state changes initiated by the call are finished before delivering the result as response. However, in long running request processing situations you may make use of asynchronous interface design with multiple calls: one for initiating the asynchronous processing and subsequent ones for accessing the processing status and/or result.

We recommend an API design that represents the asynchronous request processing explicitly via a job resource that has a status and is different from the actual business resource. For instance, `POST /report-jobs` returns HTTP status code **201** to indicate successful initiation of asynchronous processing together with the *job-id* passed in the response payload and/or via the

URL of the `Location` header. The *job-id* or `Location` URL then can be used to poll the processing status via `GET /report-jobs/ id` which returns HTTP status code **200** with job status and optional report-id as response payload. Once returned with job status `finished`, the report-id is provided and can be used to fetch the result via `GET /reports/ id` which returns **200** and the report object as response payload.

Alternatively, if you do not to follow the recommended practice of providing a separate job resource, you may use `POST /reports` returning a status code **202** together with the `Location` header to indicate successful initiation of the asynchronous processing. The `Location` URL is used to fetch the report via `GET /reports/ id` which returns either **200** and the report resource or **202** without payload, if the asynchronous processing is still ongoing.

Hint: Do **not** use response code **204** or **404** instead of **202** here — it is misleading since neither is the processing successfully finished, nor do we want to suggest a client failure.

MUST define collection format of header and query parameters [154]

Header and query parameters allow to provide a collection of values, either by providing a comma-separated list of values or by repeating the parameter multiple times with different values as follows:

Parameter Type	Comma-separated Values	Multiple Parameters	Standard
Header	`Header: value1,value2`	`Header: value1, Header: value2`	RFC 9110 Section 5.3
Query	`?param=value1,value2`	`?param=value1&param=value2`	RFC 6570 Section 3.2.8

As OpenAPI does not support both schemas at once, an API specification must explicitly define the collection format to guide consumers as follows:

Parameter Type	Comma-separated Values	Multiple Parameters
Header	`style: simple, explode: false`	not allowed (see RFC 9110 Section 5.3)

Parameter Type	Comma-separated Values	Multiple Parameters
Query	`style: form, explode: false`	`style: form, explode: true`

When choosing the collection format, take into account the tool support, the escaping of special characters and the maximal URL length.

SHOULD design simple query languages using query parameters [236]

We prefer the use of query parameters to describe resource-specific query languages for the majority of APIs because it's native to HTTP, easy to extend and has an excellent implementation support in HTTP clients and web frameworks.

By simple query language we mean one or more name-value pairs that are combined in one way only with `and` semantics.

Query parameters should have the following aspects specified:

- Reference to corresponding property, if any
- Value range, e.g. inclusive vs. exclusive
- Comparison semantics (equals, less than, greater than, etc)
- Implications when combined with other queries, e.g. *and* vs. *or*

How query parameters are named and used is up to individual API designers, here are a few tips that could help to decide whether to use simple or more complex query language:

1. Consider using simple query language when API is built to be used by others (external teams):
    - no additional effort/logic to form the query
    - no ambiguity in meaning of the query parameters. For example in `GET /items?user_id=gt:100`, is `user_id` greater than `100` or is `user_id` equal to `gt:100`?
    - easy to read, no learning curve
2. For internal usage or specific use case a more complex query language can be used (such as `price gt 10` or `price[gt]=10` or `price>10` etc.). Also please consider following our

guidance for designing complex query languages with JSON.

The following examples should serve as ideas for simple query language:

Equals

- `name=Zalando`, `creation_year=2023`, `updated_by=user1` (query elements based on property equality)
- `created_at=2023-09-18T12:12:00.000Z`, `age=18` (query elements based on logical properties)
- `color=red,green,blue,multicolored` (query elements based on multiple choice possibility)
  - for these type of filters, consider to use guidance to have smth like `filters={"color":["red","green","blue"]}`.

Less than

- `max_length=5` — query elements based on upper/lower bounds (`min` and `max`)
- `shorter_than=5` — query elements using terminology specific e.g. to *length*
- `price_lower_than=50` or `price_lower_than_or_equal=50`
- `created_before=2019-07-17` or `active_until=2023-09-18T12:12:00.000Z`
  - Using terminology specific to time: *before*, *after*, *since* and *until*

More than

- `min_length=2` — query elements based on upper/lower bounds (`min` and `max`)
- `created_after=2019-07-17` or `modified_since=2019-07-17`
  - Using terminology specific to time: *before*, *after*, *since* and *until*
- `price_higher_than=50` or `price_equal_or_higher_than=50`

Pagination

- `offset=10` and `limit=5` (query elements for pagination regardless customer sorting)
- `limit=5` and `created_after=2019-07-17` (query elements for keyset pagination)
  - when sorting is in place and new elements are inserted, it prevents showing repeated/missing results due to offset shift.

Please check conventional query parameters for pagination and sorting and you can also find additional info in REST Design - Pagination section below.

We don't advocate for or against certain names because in the end APIs should be free to choose the terminology that fits their domain the best.

SHOULD design complex query languages using JSON [237]

Minimalistic query languages based on query parameters are suitable for simple use cases with a small set of available filters that are combined in one way and one way only (e.g. *and* semantics). Simple query languages are generally preferred over complex ones.

Some APIs will have a need for sophisticated and more complex query languages. Dominant examples are APIs around search (incl. faceting) and product catalogs.

Aspects that set those APIs apart from the rest include but are not limited to:

- Unusual high number of available filters
- Dynamic filters, due to a dynamic and extensible resource model
- Free choice of operators, e.g. `and` , `or` and `not`

APIs that qualify for a specific, complex query language are encouraged to use nested JSON data structures and define them using OpenAPI directly. The provides the following benefits:

- Data structures are easy to use for clients
  - No special library support necessary
  - No need for string concatenation or manual escaping
- Data structures are easy to use for servers
  - No special tokenizers needed
  - Semantics are attached to data structures rather than text tokens
- Consistent with other HTTP methods
- API is defined in OpenAPI completely
  - No external documents or grammars needed
  - Existing means are familiar to everyone

JSON-specific rules and most certainly needs to make use of the `GET` -with-body pattern.

Example

The following JSON document should serve as an idea how a structured query might look like.

```
{
  "and": {
    "name": {
      "match": "Alice"
    },
    "age": {
      "or": {
        "range": {
          ">": 25,
          "<=": 50
        },
        "=": 65
      }
    }
  }
}
```

Feel free to also get some inspiration from:

- Elastic Search: Query DSL
- GraphQL: Queries

MUST document implicit response filtering [226]

Sometimes certain collection resources or queries will not list all the possible elements they have, but only those for which the current client is authorized to access.

Implicit filtering could be done on:

- the collection of resources being returned on a `GET` request
- the fields returned for the detail information of the resource

In such cases, the fact that implicit filtering is applied must be documented in the API specification's endpoint description. Consider caching aspects when implicit filtering is provided. Example:

If an employee of the company *Foo* accesses one of our business-to-business service and

performs a `GET /business-partners`, it must, for legal reasons, not display any other business partner that is not owned or contractually managed by her/his company. It should never see that we are doing business also with company *Bar*.

Response as seen from a consumer working at `FOO`:

```
{
    "items": [
        { "name": "Foo Performance" },
        { "name": "Foo Sport" },
        { "name": "Foo Signature" }
    ]
}
```

Response as seen from a consumer working at `BAR`:

```
{
    "items": [
        { "name": "Bar Classics" },
        { "name": "Bar pour Elle" }
    ]
}
```

The API Specification should then specify something like this:

```
paths:
  /business-partner:
    get:
      description: >-
        Get the list of registered business partner.
        Only the business partners to which you have access to are returned.
```

---

10. REST Basics - HTTP status codes

MUST use official HTTP status codes [243]

You must only use official HTTP status codes consistently with their intended semantics. Official HTTP status codes are defined via RFC standards and registered in the IANA Status Code Registry. Main RFC standards are RFC 9110 - HTTP Semantics, section 15: Status Codes and RFC 6585 - HTTP: Additional Status Codes. Wikipedia: HTTP status codes provides an overview on the official error codes (which also lists some unofficial status codes, e.g. defined by popular

web servers like Nginx, that we do not suggest to use).

MUST specify success and error responses [151]

APIs should define the functional, business view and abstract from implementation aspects. Success and error responses are a vital part to define how an API is used correctly.

Therefore, you must define **all** success and service specific error responses in your API specification. Both are part of the interface definition and provide important information for service clients to handle standard as well as exceptional situations. Error code response descriptions should provide information about the specific conditions that lead to the error, especially if these conditions can be changed by how the endpoint is used by the clients.

API designers should also think about a **troubleshooting board** as part of the associated online API documentation. It provides information and handling guidance on application-specific errors and is referenced via links from the API specification. This can reduce service support tasks and contribute to service client and provider performance.

Exception: Standard client and server errors, e.g. **401** (unauthenticated), **403** (unauthorized), **404** (not found), **500** (internal server error), or **503** (service unavailable), where the semantic can be easily derived from the end endpoint specification need no individual definition. Instead these can be included by applying the `default` shown pattern below. However, error codes that provide endpoint specific indications for clients on how to avoid calling the endpoint in the wrong way, or be prepared to react on specific error situation must be specified explicitly.

```
responses:
  ...
  default:
    description: error occurred - see status code and problem object for more information.
    content:
      "application/problem+json":
        schema:
          $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/problem-1.0.1.yaml#/Problem'
```

SHOULD only use most common HTTP status codes [150]

The most commonly used codes are best understood and listed below as subset of the official HTTP status codes and consistent with their semantics in the RFCs. We avoid less commonly used codes that easily create misconceptions due to less familiar semantics and API specific interpretations.

Important: As long as your HTTP status code usage is well covered by the semantic defined here, you should not describe it to avoid an overload with common sense information and the risk of inconsistent definitions. Only if the HTTP status code is not in the list below or its usage requires additional information aside the well defined semantic, the API specification must provide a clear description of the HTTP status code in the response.

Legend

A more specific version of the Conventions used in these guidelines.

✔ use

A common, well understood status code that should be used where appropriate. Does NOT mean that every operation must be able to return this code.

✖ do not use

We do not see a good use-case for returning this status code from a RESTful API. The status code might be applicable in other contexts, e.g. returned by reverse proxies, for web pages, etc. Implicitly also means ✖ do not document, as status codes that are not returned by the API should also not be documented.

✔ document

If the status code can be returned by the API, it should be documented in the API specification.

✖ do not document

The status code has a well-understood standard meaning to it, so only document it if there are operation specific details you want to add. See exception in rule 151.

Success codes

200 OK RFC ✔ use ✔ document `<all>`

This is the most general success response. It should be used if the more specific codes below are not applicable. In a robust resource creation via `POST`, `PUT`, and `PATCH` in conjunction with a **201** this indicates that the returned resource exists before.

201 Created RFC ✔ use ✔ document `POST` `PUT`

Returned on successful resource creation. **201** is returned with or without response payload (unlike **200** / **204**). We recommend additionally to return the created resource

URL via the `Location` response header (see **MAY** use standard headers).

202 Accepted RFC ✔ use ✔ document `POST` `PUT` `PATCH` `DELETE`

The request was successful and will be processed asynchronously. Only applicable to methods which change something, with the exception of `GET` methods indicating that a resources is still created asynchronously as described in **MAY** support asynchronous request processing.

204 No content RFC ✔ use ✔ document `POST` `PUT` `PATCH` `DELETE`

Returned instead of **200**, if no response payload is returned. Normally only applicable to methods which change something.

205 Reset Content RFC ✖ do not use `<all>`

This is meant for interactive use cases, e.g. to clear a form after submitting it. There is no reason to use it in a REST API.

206 Partial Content RFC ✖ do not use `<all>`

For responses to range requests, when only a part of the resource indicated by the byte range is returned. This is not for pagination, where a normal **200** should be used. This might be useful in rare cases (like media streaming or downloading large files), but most APIs don't need this.

207 Multi-Status RFC ✔ use ✔ document `POST` `(DELETE)`

The response body contains status information for multiple different parts of a batch/bulk request (see **MUST** use code 207 for batch or bulk requests for details). Normally used for `POST`, in some cases also for `DELETE`.

Redirection codes

301 Moved Permanently RFC ✖ do not use `<all>`

This and all future requests should be directed to the given URI. See also **SHOULD** not use redirection codes.

302 Found RFC ✖ do not use `<all>`

This is a temporary redirect where some clients **MAY** change the request method from `POST` to `GET`. Mainly used for dismissing and redirecting form submissions in

browsers. See also **SHOULD** not use redirection codes.

303 See Other RFC ✖ do not use `POST` `PUT` `PATCH` `DELETE`

The response to the request can be found under another URI using a `GET` method. A disambiguated version of **302** for the case where the client **MUST** change the method to `GET`. See also **SHOULD** not use redirection codes.

304 Not Modified RFC ✔ document `GET` `HEAD`

Indicates that a conditional `GET` or `HEAD` request would have resulted in **200** response, if it were not for the fact that the condition evaluated to false, i.e. resource has not been modified since the date or version passed via request headers `If-Modified-Since` or `If-None-Match`. For `PUT` / `PATCH` / `DELETE` requests, use **412** instead.

307 Temporary Redirect RFC ✖ do not use `<all>`

The response to the request can be found under another URI. A disambiguated version of **302** where the client **MUST** keep the same method as the original request. See also **SHOULD** not use redirection codes.

308 Permanent Redirect RFC ✖ do not use `<all>`

Similar to **307**, but the client should persist the new URI. Applicable more to browsers. For APIs, the URI should be explicitly fixed at the source instead of being implicitly kept in some state based on a previous redirect. See also **SHOULD** not use redirection codes.

Client side error codes

400 Bad Request RFC ✔ use ✔ document `<all>`

Unspecific client error indicating that the server cannot process the request due to something that is perceived to be a client error (e.g. malformed request syntax, invalid request). Should also be delivered in case of input payload fails business logic / semantic validation (instead of using **422**).

401 Unauthorized RFC ✔ use ✖ do not document `<all>`

Actually **Unauthenticated**. The credentials are missing or not valid for the target resource. For an API, this usually means that the provided token or cookie is not valid. As this can happen for almost every endpoint, APIs should normally not document this.

403 Forbidden RFC ✖ do not document `<all>`

The user is not authorized to use this resource. For an API, this can mean that the request's token was valid, but was missing a scope for this endpoint. Or that some object-specific authorization failed. We recommend only documenting the second case.

404 Not found RFC ✔ use ✖ do not document `<all>`

The target resource was not found. This will be returned by most paths on most APIs (with out being documented), and for endpoints with parameters when those parameters cannot be map to an existing entity. For a `PUT` endpoint which only supports updating existing resources, this might be returned if the resource does not exist. Apart from these special cases, this does not need to be documented.

405 Method Not Allowed RFC ✔ document `<all>`

The request method is not supported for this resource. In theory, this can be returned for all resources for all the methods except the ones documented. Using this response code for an existing endpoint (usually with path parameters) only makes sense if it depends on some internal resource state whether a specific method is allowed, e.g. an order can only be canceled via `DELETE` until the shipment leaves the warehouse. **Do not use it unless you have such a special use case, but then make sure to document it, making it clear why a resource might not support a method.**

406 Not Acceptable RFC ✔ use ✖ do not document `<all>`

Resource only supports generating content with content-types that are not listed in the `Accept` header sent in the request.

408 Request timeout RFC ✖ do not use `<all>`

The server times out waiting for the request to arrive. For APIs, this should not be used.

409 Conflict RFC ✔ document `POST` `PUT` `PATCH` `DELETE`

The request cannot be completed due to conflict with the current state of the target resource. For example, you may get a **409** response when updating a resource that is older than the existing one on the server, resulting in a version control conflict. If this is used, it **MUST** be documented. For successful robust creation of resources (`PUT` or `POST`) you should always return **200** or **204** and not **409**, even if the resource exists already. If any `If-*` headers cause a conflict, you should use **412** and not **409**. Only applicable to methods which change something.

410 Gone RFC ✖ do not document `<all>`

The resource does not exist any longer. It did exist in the past, and will most likely not exist in the future. This can be used, e.g. when accessing a resource that has intentionally been deleted. This normally does not need to be documented, unless there is a specific need to distinguish this case from the normal **404**.

411 Length Required RFC ✔ document `POST` `PUT` `PATCH`

The server requires a `Content-Length` header for this request. This is normally only relevant for large media uploads. The corresponding header parameter should be marked as required. If used, it **MUST** to be documented (and explained).

412 Precondition Failed RFC ✔ use ✖ do not document `PUT` `PATCH` `DELETE`

Returned for conditional requests, e.g. `If-Match` if the condition failed. Used for optimistic locking. Normally only applicable to methods that change something. For `HEAD` / `GET` requests, use **304** instead.

415 Unsupported Media Type RFC ✔ use ✖ do not document `POST` `PUT` `PATCH`

The client did not provide a supported content-type for the request body. Only applicable to methods with a request body.

417 Expectation Failed RFC ✖ do not use `<all>`

Returned when the client used an `Expect` header which the server does not support. The only defined value for the `Expect` header is very technical and does not belong in an API.

418 I'm a teapot 🫖 (Unused) RFC ✖ do not use `<all>`

Only use if you are implementing an API for a teapot that does not support brewing coffee. Response defined for April's Fools in RFC 2324.

422 Unprocessable Content RFC ✖ do not use `<all>`

The server understands the content type, but is unable to process the content. We do not recommend this code to be used as **400** already covers most use-cases and there does not seem to be a clear benefit to differentiating between them.

423 Locked RFC ✔ document `PUT` `PATCH` `DELETE`

Pessimistic locking, e.g. processing states. May be used to indicate an existing resource

lock, however, we recommend using optimistic locking instead. If used, it must be documented to indicate pessimistic locking.

424 Failed Dependency RFC ✖ do not use `<all>`

The request failed due to failure of a previous request. This is not applicable to restful APIs.

428 Precondition Required RFC ✔ use ✖ do not document `<all>`

Server requires the request to be conditional, e.g. to make sure that the "lost update problem" is avoided (see **MAY** consider to support `ETag` together with `If-Match` / `If-None-Match` header). Instead of documenting this response status, the required headers should be documented (and marked as required).

429 Too many requests RFC ✔ use ✖ do not document `<all>`

The client is not abiding by the rate limits in place and has sent too many requests (see **MUST** use code 429 with headers for rate limits).

431 Request Header Fields Too Large RFC ✖ do not document `<all>`

The server is not able to process the request because the request headers are too large. Usually used by gateways and proxies with memory limits.

Server side error codes

500 Internal Server Error RFC ✔ use ✖ do not document `<all>`

A generic error indication for an unexpected server execution problem. Clients should be careful with retrying on this response, since the nature of the problem is unknown and must be expected to continue.

501 Not Implemented RFC ✔ document `<all>`

Server cannot fulfill the request, since the endpoint is not implemented yet. Usually this implies future availability, but retrying now is not recommended. May be documented on endpoints that are planned to be implemented in the future to indicate that they are still not available.

502 Bad Gateway RFC ✔ use ✖ do not document `<all>`

The server, while acting as a gateway or proxy, received an invalid response from an

inbound server attempting to fulfill the request. May be used by a server to indicate that an inbound service is creating an unexpected result instead of **500**. Clients should be careful with retrying on this response, since the nature of the problem is unknown and must be expected to continue.

503 Service Unavailable RFC ✔ use ✖ do not document `<all>`

Service is (temporarily) not available, e.g. if a required component or inbound service is not available. Client are encouraged to retry requests following an exponential back off pattern. If possible, the service should indicate how long the client should wait by setting the `Retry-After` header.

504 Gateway Timeout RFC ✔ use ✖ do not document `<all>`

The server, while acting as a gateway or proxy, did not receive a timely response. May be used by servers to indicate that an inbound service cannot process the request fast enough. Client may retry the request immediately exactly once, to check whether warming up the service solved the problem.

505 HTTP Version Not Supported RFC ✖ do not use `<all>`

The server does not support the HTTP protocol version used in the request. Technical response code that serves not use case in RESTful APIs.

507 Insufficient Storage RFC ✖ do not document `POST` `PUT` `PATCH`

The server is unable to store the resource as needed to complete the request. May be used to indicate that the server is out of disk space.

511 Network Authentication Required RFC ✖ do not use `<all>`

The client needs to authenticate to gain network access. Technical response code that serves no use case in RESTful APIs.

MUST use most specific HTTP status codes [220]

You must use the most specific HTTP status code when returning information about your request processing status or error situations.

MUST use code 207 for batch or bulk requests [152]

Some APIs are required to provide either *batch* or *bulk* requests using `POST` for performance

reasons, i.e. for communication and processing efficiency. In this case services may be in need to signal multiple response codes for each part of a batch or bulk request. As HTTP does not provide proper guidance for handling batch/bulk requests and responses, we herewith define the following approach:

- A batch or bulk request **always** responds with HTTP status code **207** unless a non-item-specific failure occurs.
- A batch or bulk request **may** return **4xx**/**5xx** status codes, if the failure is non-item-specific and cannot be restricted to individual items of the batch or bulk request, e.g. in case of overload situations or general service failures.
- A batch or bulk response with status code **207 always** returns as payload a multi-status response containing item specific status and/or monitoring information for each part of the batch or bulk request.

Note: These rules apply *even in the case* that processing of all individual parts *fail* or each part is executed *asynchronously*!

The rules are intended to allow clients to act on batch and bulk responses in a consistent way by inspecting the individual results. We explicitly reject the option to apply **200** for a completely successful batch as proposed in Nakadi's `POST /event-types/{name}/events` as short cut without inspecting the result, as we want to avoid risks and expect clients to handle partial batch failures anyway.

The bulk or batch response may look as follows:

```
BatchOrBulkResponse:
  description: batch response object.
  type: object
  properties:
    items:
      type: array
      items:
        type: object
        properties:
          id:
            description: Identifier of batch or bulk request item.
            type: string
          status:
            description: >
              Response status value. A number or extensible enum describing
              the execution status of the batch or bulk request items.
            type: string
```

```
          x-extensible-enum: [...]
        description:
          description: >
            Human readable status description and containing additional
            context information about failures etc.
          type: string
      required: [id, status]
```

Note: while a *batch* defines a collection of requests triggering independent processes, a *bulk* defines a collection of independent resources created or updated together in one request. With respect to response processing this distinction normally does not matter.

MUST use code 429 with headers for rate limits [153]

APIs that wish to manage the request rate of clients must use the **429** (Too Many Requests) response code, if the client exceeded the request rate (see RFC 6585). Such responses must also contain header information providing further details to the client. There are two approaches a service can take for header information:

- Return a `Retry-After` header indicating how long the client ought to wait before making a follow-up request. The Retry-After header can contain a HTTP date value to retry after or the number of seconds to delay. Either is acceptable but APIs should prefer to use a delay in seconds.
- Return a trio of `X-RateLimit` headers. These headers (described below) allow a server to express a service level in the form of a number of allowing requests within a given window of time and when the window is reset.

The `X-RateLimit` headers are:

- `X-RateLimit-Limit` : The maximum number of requests that the client is allowed to make in this window.
- `X-RateLimit-Remaining` : The number of requests allowed in the current window.
- `X-RateLimit-Reset` : The relative time in seconds when the rate limit window will be reset. **Beware** that this is different to Github and Twitter's usage of a header with the same name which is using UTC epoch seconds instead.

The reason to allow both approaches is that APIs can have different needs. Retry-After is often sufficient for general load handling and request throttling scenarios and notably, does not strictly require the concept of a calling entity such as a tenant or named account. In turn this allows

resource owners to minimize the amount of state they have to carry with respect to client requests. The 'X-RateLimit' headers are suitable for scenarios where clients are associated with pre-existing account or tenancy structures. 'X-RateLimit' headers are generally returned on every request and not just on a 429, which implies the service implementing the API is carrying sufficient state to track the number of requests made within a given window for each named entity.

MUST support problem JSON [176]

RFC 9457 defines a Problem JSON object using the media type `application/problem+json` to provide an extensible human and machine readable failure information beyond the HTTP response status code to transports the failure kind (`type` / `title`) and the failure cause and location (`instant` / `detail`). To make best use of this additional failure information, every endpoints must be capable of returning a Problem JSON on client usage errors (**4xx** status codes) as well as server side processing errors (**5xx** status codes).

Note: Clients must be robust and **not rely** on a Problem JSON object being returned, since (a) failure responses may be created by infrastructure components not aware of this guideline or (b) service may be unable to comply with this guideline in certain error situations.

Hint: The media type `application/problem+json` is often not implemented as a subset of `application/json` by libraries and services! Thus clients need to include `application/problem+json` in the `Accept` -Header to trigger delivery of the extended failure information.

The OpenAPI schema definition of the Problem JSON object can be found on GitHub. You can reference it by using:

```yaml
responses:
  503:
    description: Service Unavailable
    content:
      "application/problem+json":
        schema:
          $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/problem-1.0.1.yaml#/Problem'
```

You may define custom problem types as extensions of the Problem JSON object if your API needs to return specific, additional, and more detailed error information.

Note: Problem `type` and `instance` identifiers in our APIs are not meant to be resolved. RFC 9457 encourages that problem types are URI references that point to human-readable

documentation, **but** we deliberately decided against that, as all important parts of the API must be documented using OpenAPI anyway. In addition, URLs tend to be fragile and not very stable over longer periods because of organizational and documentation changes and descriptions might easily get out of sync.

In order to stay compatible with RFC 9457 we proposed to use relative URI references usually defined by `absolute-path [ '?' query ] [ '#' fragment ]` as simplified identifiers in `type` and `instance` fields:

- `/problems/out-of-stock`
- `/problems/insufficient-funds`
- `/problems/user-deactivated`
- `/problems/connection-error#read-timeout`

Hint: The use of absolute URIs is not forbidden but strongly discouraged. If you use absolute URIs, please reference problem-1.0.0.yaml#/Problem instead.

MUST not expose stack traces [177]

Stack traces contain implementation details that are not part of an API, and on which clients should never rely. Moreover, stack traces can leak sensitive information that partners and third parties are not allowed to receive and may disclose insights about vulnerabilities to attackers.

SHOULD not use redirection codes [251]

We generally do not recommend using redirection codes for most API cases (except for **304**, which is not really a redirection code). Usually you would use the redirection to migrate clients to a new service location. However, this is better accomplished by one of the following.

1. Changing the clients to use the new location in the first place, avoiding the need for redirection.
2. Redirecting the traffic behind the API layer (e.g. in the reverse proxy or the app itself) without the client having to be involved.
3. Deprecating the endpoint and removing it as described in REST Design - Deprecation.

For idempotent `POST` cases, where you want to inform the client that a resource already exists at a certain location, you should instead use **200** with the `Location` header set. This is along the

same lines as the creation case where **201** is used instead. See also **SHOULD** consider to design `POST` and `PATCH` idempotent.

For non-idempotent `POST` cases, where you want to inform the client that the resource has already been created and cannot be created again (e.g. payment), you should return **409** instead of redirecting to make the error case more explicit.

---

11. REST Basics - HTTP headers

In this section we explain a handful of (standard) HTTP headers, which we found most useful in certain situations and that require additional explanation to be used effectively.

Note: we generally discourage the usage of proprietary headers. However, they are sometimes useful to pass generic, service independent, overarching information relevant for our specific application architecture. We consistently define these proprietary headers below.

Whether a service supports a certain header is part of the service contract. Therefore APIs should use the `parameters` and `headers` definition of the API endpoint and response to clarify what is supported in a certain context.

Using Standard Header definitions

Usually, you can the standard HTTP request and response header definition provided by the guideline to simplify API by using well recognized patterns. The best practice importing headers providing the highest readability is as follows:

```
path:
  '/resource'
    get:
      parameters:
        - $ref: '#/components/parameters/ETag

components:
  parameters|headers:
    ETag:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/ETag'

  responses:
    Default:
      headers:
        ETag:
```

```
        $ref: '#/components/(parameters|headers)/ETag'
```

Note: It is a question of taste whether headers for responses are defined in `#/components/headers` or `#/components/parameters`. Unfortunately, headers in the first section cannot be referenced from a `parameters`-list, while it is possible to reference the second also from a `headers`-list.

MAY use standard headers [133]

APIs may make use of HTTP headers defined by non-obsolete RFCs (see following list of standard HTTP headers). The supported headers must be explicitly mentioned in the API specification.

SHOULD use kebab-case with uppercase separate words for HTTP headers [132]

This convention is followed by (most of) the standard headers e.g. as defined in RFC 2616, RFC 4229, RFC 9110. Examples:

```
If-Modified-Since
Accept-Encoding
Content-ID
Language
```

Note: The HTTP standard defines headers (now called HTTP fields) to be case-insensitive (see RFC 9110 Section 5.1). However, for the sake of readability and consistency, you should follow the convention when defining standard or proprietary headers. Exceptions are common abbreviations like `ID`.

MUST use `Content-*` headers correctly [178]

Content or entity headers are headers with a `Content-` prefix. They describe the content of the body of the message and they can be used in both, HTTP requests and responses. Commonly used content headers include but are not limited to:

- `Content-Disposition` can indicate that the representation is supposed to be saved as a file, and the proposed file name.
- `Content-Encoding` indicates compression or encryption algorithms applied to the content.
- `Content-Length` indicates the length of the content (in bytes).

- `Content-Language` indicates that the body is meant for people literate in some human language(s).
- `Content-Location` indicates where the body can be found otherwise (**MAY** use `Content-Location` header for more details]).
- `Content-Range` is used in responses to range requests to indicate which part of the requested resource representation is delivered with the body.
- `Content-Type` indicates the media type of the body content.

SHOULD use `Location` header instead of `Content-Location` header [180]

As a correct usage of the `Content-Location` response header (see **MAY** use `Content-Location` header) with respect to caching and its method specific semantics is difficult, we **discourage** the use of `Content-Location`. In most cases it is sufficient to inform clients about the resource location in create or re-direct responses by using the `Location` header while avoiding the `Content-Location` specific ambiguities and complexities.

For more details see RFC 9110 Section 10.2.2 Location, RFC 9110 Section 8.2 Content-Location

MAY use `Content-Location` header [179]

The `Content-Location` response header is an *optional* header that can be used in successful write operations (`PUT`, `POST`, or `PATCH`) and read operations (`GET`, `HEAD`) to guide caching and signal a receiver the actual location of the resource transmitted in the response body. This allows clients to identify the resource and to update their local copy when receiving a response with this header.

The `Content-Location` header can be used to support the following use cases:

- For reading operations `GET` and `HEAD`, a different location than the requested URL can be used to indicate that the returned resource is subject to content negotiations, and that the value provides a more specific identifier of the resource.
- For writing operations `PUT` and `PATCH`, an identical location to the requested URL can be used to explicitly indicate that the returned resource is the current representation of the newly created or updated resource.
- For writing operations `POST` and `DELETE`, a content location can be used to indicate that the body contains a status report resource in response to the requested action, which is available

at provided location.

Note: When using the `Content-Location` header, the `Content-Type` header has to be set as well. For example:

```
GET /products/123/images HTTP/1.1

HTTP/1.1 200 OK
Content-Type: image/png
Content-Location: /products/123/images?format=raw
```

See also **MUST** document cacheable `GET`, `HEAD`, and `POST` endpoints.

MAY consider to support `Prefer` header to handle processing preferences [181]

The `Prefer` header defined in RFC 7240 allows clients to request processing behaviors from servers. It pre-defines a number of preferences and is extensible, to allow others to be defined. Support for the `Prefer` header is entirely optional and at the discretion of API designers, but as an existing Internet Standard, is recommended over defining proprietary "X-" headers for processing directives.

The `Prefer` header can be defined in the API specification as follows:

```yaml
components:
  parameters:
    Prefer:
        # Do not import this schema directly, since processing directives are usually
        # highly customized. Instead, copy the schema to your API and adjust it to
        # your needs.
        name: Prefer
        in: header
        required: false
        description: |-
          The **Prefer** header indicates that a particular server behavior is
          preferred by the client, but is not required for successful completion of
          the request (see [RFC 7240][rfc-7240]. The following behaviors are
          supported by this API:

          * **respond-async** is used to suggest the server to respond as fast as
            possible asynchronously using **202** (Accepted) instead of waiting for
            the result.
          * **return=<minimal|representation>** is used to suggest the server to
            return using **204** (No Content) without resource (minimal) or using
            **200** or **201** with resource (representation) in the response body on
```

```
      success.
    * **return=<total-count>** is used to suggest the server to return a total
      result count in a collection requests supporting pagination. Since this
      is a costly operation, it should be used with care, and the service may
      decide to ignore this request.
    * **wait=<delta-seconds>** is used to suggest a maximum time the server has
      time to process the request synchronously.
    * **handling=<strict|lenient>** is used to suggest the server to be strict
      and report error conditions or lenient, i.e. robust and try to continue,
      if possible.

    See [RFC 7240][rfc-7240] as well as [API Guideline Rule #181][api-181]
    for further details.

    [rfc-7240]: <https://tools.ietf.org/html/rfc7240>
    [api-181]: <https://opensource.zalando.com/restful-api-guidelines/#181>
  schema:
    type: array
    items:
      type: string
  style: simple
  explode: false
  example: [ "respond-async", "return=minimal", "wait=20", "handling=strict",
    "respond-async,return=minimal,wait=20,handling=strict" ]
```

Note: Please, copy only the behaviors into your `Prefer` header specification that are supported by your API endpoint. If necessary, specify different `Prefer` headers for each supported use case.

Supporting APIs may return the `Preference-Applied` header also defined in RFC 7240 to indicate whether a preference has been applied.

MAY consider to support `ETag` together with `If-Match` / `If-None-Match` header [182]

When creating or updating resources it may be necessary to expose conflicts and to prevent the *lost update* or *initially created* problem. Following RFC 9110 Section 13 "Conditional Requests" this can be best accomplished by supporting the `ETag` header together with the `If-Match` or `If-None-Match` conditional header. The contents of an `ETag: <entity-tag>` header is either (a) a hash of the response body, (b) a hash of the last modified field of the entity, or (c) a version number or identifier of the entity version.

To expose conflicts between concurrent update operations via `PUT`, `POST`, or `PATCH`, the `If-Match: <entity-tag>` header can be used to force the server to check whether the version of the updated entity is conforming to the requested `<entity-tag>`. If no matching entity is found, the

operation is supposed a to respond with status code **412** - precondition failed.

Beside other use cases, `If-None-Match: *` can be used in a similar way to expose conflicts in resource creation. If any matching entity is found, the operation is supposed a to respond with status code **412** - precondition failed.

The `ETag`, `If-Match`, and `If-None-Match` headers can be defined as follows in the API specification (see also the default definition below):

```yaml
components:
  parameters|headers:
    ETag:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/ETag'
    If-Match:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/If-Match'
    If-None-Match:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/If-None-Match'
```

```yaml
components:
  headers:
    ETag:
        required: false
        description: |-
          The **ETag** header field in a response provides an opaque quoted string
          identifying the distinct delivered resource. The same selected resource
          depending on version and representation may be identified by multiple
          identifiers. The **ETag** value is guaranteed to change whenever the
          resource changes, and thereby enabling optimistic updates.

          An identifier consists of an opaque quoted string, possibly prefixed by
          a weakness indicator. See [RFC 9110 Section 8.8.3][rfc-9110-8.8.3] as
          well as [API Guideline Rule #182][api-182] for further details.

          [rfc-9110-8.8.3]: <https://tools.ietf.org/html/rfc9110#section-8.8.3>
          [api-182]: <https://opensource.zalando.com/restful-api-guidelines/#182>
        schema:
          type: array
          items:
            type: string
        style: simple
        explode: false
        example: [ W/"xy", "5", "5db68c06-1a68-11e9-8341-68f728c1ba70" ]

    If-Match:
        name: If-Match
        in: header
        required: false
        description: |-
```

```
        The **If-Match** header field is used to declare a list of identifiers that
        are required to match the current resource version identifier in at least
        one position as a pre-condition for executing the request on the server
        side. This behavior is used to validate and reject optimistic updates, by
        checking if the resource version a consumer has based his changes on is
        outdated on arrival of the change request to prevent lost updates.

        If the pre-condition fails the server will respond with status code **412**
        (Precondition Failed). See [RFC 9110 Section 13.1.1][rfc-9110-13.1.1] as
        well as [API Guideline Rule #182][api-182] for further details.

        [rfc-9110-13.1.1]: <https://tools.ietf.org/html/rfc9110#section-13.1.1>
        [api-182]: <https://opensource.zalando.com/restful-api-guidelines/#182>
      schema:
        type: array
        items:
          type: string
      style: simple
      explode: false
      example: [ "5db68c06-1a68-11e9-8341-68f728c1ba70", 'W/"xy", "5"' ]

  If-None-Match:
      name: If-None-Match
      in: header
      required: false
      description: |-
        The **If-None-Match header** field is used to declare a list of identifiers
        that are required to fail matching all the current resource version
        identifiers as a pre-condition for executing the request on the server
        side. This is especially used in conjunction with an **\*** (asterix) that
        is matching all possible resource identifiers to ensure the initial
        creation of a resource. Other use cases are possible but rare.

        If the pre-condition fails the server will respond with status code **412**
        (Precondition Failed). See [RFC 9110 Section 13.1.2][rfc-9110-13.1.2] as
        well as [API Guideline Rule #182][api-182] for further details.

        [rfc-9110-13.1.2]: <https://tools.ietf.org/html/rfc9110#section-13.1.2>
        [api-182]: <https://opensource.zalando.com/restful-api-guidelines/#182>
      schema:
        type: array
        items:
          type: string
      style: simple
      explode: false
      example: [ "5db68c06-1a68-11e9-8341-68f728c1ba70", 'W/"xy", "5"' ]
```

Please see Optimistic locking in RESTful APIs for a detailed discussion as well as Cache Support Patterns for additional use cases.

MAY consider to support `Idempotency-Key` header [230]

When creating or updating resources it can be helpful or necessary to ensure a strong idempotent behavior comprising same responses, to prevent duplicate execution in case of retries after timeout and network outages. Generally, this can be achieved by sending a client specific *unique request key* – that is not part of the resource – via `Idempotency-Key` header.

The *unique request key* is stored temporarily, e.g. for 24 hours, together with the response and the request hash (optionally) of the first request in a key cache, regardless of whether it succeeded or failed. The service can now look up the *unique request key* in the key cache and serve the response from the key cache, instead of re-executing the request, to ensure idempotent behavior. Optionally, it can check the request hash for consistency before serving the response. If the key is not in the key store, the request is executed as usual and the response is stored in the key cache.

This allows clients to safely retry requests after timeouts, network outages, etc. while receive the same response multiple times. **Note:** The request retry in this context requires to send the exact same request, i.e. updates of the request that would change the result are off-limits. The request hash in the key cache can protection against this misbehavior. The service is recommended to reject such a request using status code **400**.

Important: To grant a reliable idempotent execution semantic, the resource and the key cache have to be updated with hard transaction semantics – considering all potential pitfalls of failures, timeouts, and concurrent requests in a distributed systems. This makes a correct implementation exceeding the local context very hard.

The `Idempotency-Key` header must be defined as follows, but you are free to choose your expiration time:

```
components:
  parameters:
    Idempotency-Key:
        name: Idempotency-Key
        in: header
        required: false
        description: |-
          The **Idempotency-Key** is a free identifier created by the client to
          identify a request. It is used by the service to identify repeated request
          to ensure idempotent behavior by sending the same (or a similar) response
          without executing the request a second time.

          Clients should be careful as any subsequent requests with the same key may
```

```
          return the same response without further check. Thus, it is recommended to
          use a UUID version 4 (random) or any other random string with enough
          entropy to avoid collisions.

          Keys expire after 24 hours. Clients are responsible to stay within this
          limit, if they require idempotent behavior.

          See [API Guideline Rule #181][api-230] for further details.

          [api-230]: <https://opensource.zalando.com/restful-api-guidelines/#230>
        schema:
          type: string
          format: uuid
        example: [ "7da7a728-f910-11e6-942a-68f728c1ba70" ]
```

If you do not want to change the expiration time, you can also use the standard definition provided by the guideline:

```
components:
  parameters|headers:
    Idempotency-Key:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/Idempotency-Key'
```

Hint: The key cache is not intended as request log, and therefore should have a limited lifetime, else it could easily exceed the data resource in size.

Note: The `Idempotency-Key` header unlike other headers in this section is not standardized in an RFC. Our only reference are the usage in the Stripe API. However, we do not want to change the header name and semantic, and do not name it like the proprietary headers below. The header addresses a generic REST concern and is different from the Zalando landscape specific proprietary headers.

SHOULD use only the specified proprietary Zalando headers [183]

As a general rule, proprietary HTTP headers should be avoided. In addition from a conceptual point of view, the business semantics and intent of an operation should always be expressed via the path and query parameters, the method, and the content, but not via proprietary headers.

Headers are typically used to implement protocol processing aspects, such as flow control, content negotiation, and authentication, and represent business agnostic request modifiers that provide generic context information (RFC 9110 Section 10 "Message Context").

However, the exceptional usage of proprietary headers is still helpful when domain-specific

generic context information

1. needs to be passed *end-to-end* along the service call chain (even if not all called services use it as input for steering service behavior, e.g. `X-Sales-Channel` header), and/or
2. is provided by specific gateway components, for instance, our Fashion Shop API or Merchant API gateway.

Below, we explicitly define the list of proprietary headers usable for all services for passing through generic context information of our fashion domain (use case 1).

Per convention, non standardized, proprietary header names are prefixed with `X-` and use the dash (`-`) as separator (dash-case). (Due to backward compatibility, we do not follow the recommendation of Internet Engineering Task Force in RFC 6648 to deprecate the usage of `X-` `headers.) Remember that HTTP header field names are not case-sensitive:

Header field name	Type	Description	Header field value example
`X-Flow-ID`	String	For more information see **MUST** support `X-Flow-ID` .	GKY7oDhpSiKY_gAAAABZ_A
`X-Tenant-ID`	String	Identifies the tenant initiated the request to the multi tenant Zalando Platform. The `X-Tenant-ID` must be set according to the Business Partner ID extracted from the OAuth token when a request from a Business Partner hits the Zalando Platform.	9f8b3ca3-4be5-436c-a847-9cd55460c495
`X-Sales-Channel`	String	Sales channels are owned by retailers and represent a specific	52b96501-0f8d-43e7-82aa-8a96fab134d7

Header field name	Type	Description	Header field value example
		consumer segment being addressed with a specific product assortment that is offered via CFA retailer catalogs to consumers (see fashion platform glossary (internal link)).	
`X-Frontend-Type`	String	Consumer facing applications (CFAs) provide business experience to their customers via different frontend application types, for instance, mobile app or browser. Info should be passed-through as generic aspect — there are diverse concerns, e.g. pushing mobiles with specific coupons, that make use of it. Current range is mobile-app, browser, facebook-app, chat-app, email.	mobile-app
`X-Device-Type`	String	There are also use cases for steering customer experience (incl. features and content) depending on device type. Via this header info should be	tablet

Header field name	Type	Description	Header field value example
		passed-through as generic aspect. Current range is smartphone, tablet, desktop, other.	
`X-Device-OS`	String	On top of device type above, we even want to differ between device platform, e.g. smartphone Android vs. iOS. Via this header info should be passed-through as generic aspect. Current range is iOS, Android, Windows, Linux, MacOS.	Android
`X-Mobile-Advertising-ID`	String	It is either the IDFA (Apple Identifier for mobile Advertising) for iOS, or the GAID (Google mobile Advertising Identifier) for Android. It is a unique, customer-resettable identifier provided by mobile device's operating system to facilitate personalized advertising, and usually passed by mobile apps via HTTP header when calling backend services. Called	b89fadce-1f42-46aa-9c83-b7bc49e76e1f

Header field name	Type	Description	Header field value example
		services should be ready to pass this parameter through when calling other services. It is not sent if the customer disables it in the settings for respective mobile platform.	

Exception: The only exception to this guideline are the conventional hop-by-hop `X-RateLimit-` headers which can be used as defined in **MUST** use code 429 with headers for rate limits.

As part of the guidelines we provide the default definition of all proprietary headers, so you can simply reference them when defining the API endpoint. For details see Using Standard Header definitions.

Hint: This guideline does not standardize proprietary headers for our specific gateway components (2. use case above). This include, for instance, non pass-through headers `X-Zalando-Customer`, `X-Zalando-Client-ID`, `X-Zalando-Request-Host`, `X-Zalando-Request-URI` defined by Fashion Shop API (RKeep), or `X-Consumer`, `X-Consumer-Signature`, `X-Consumer-Key-ID` defined by Merchant API gateway, or `X-App-Version`, `X-Country-Code`, `X-Zalando-Auth`, `X-Forwarded-For` defined by Transactions Checkout Platform. All these proprietary headers are allow-listed in the API Linter (Zally) checking this rule.

MUST propagate proprietary headers [184]

All Zalando's proprietary headers defined in **SHOULD** use only the specified proprietary Zalando headers are end-to-end headers [2] and must be propagated to the services down the call chain. The header names and values must remain unchanged.

The values of custom headers can influence query results (e.g. `X-Device-Type` can affect recommendation results by conveying device-type information).

Sometimes the value of a proprietary header will be used as part of the entity in a subsequent request. In such cases, the proprietary headers must still be propagated as headers with the

subsequent request, despite the duplication of information.

MUST support [233] `X-Flow-ID`

The `Flow-ID` is a generic parameter to be passed through service APIs and events and written into log files and traces. A consequent usage of the `Flow-ID` facilitates the tracking of call flows through our system and allows the correlation of service activities initiated by a specific call. This is extremely helpful for operational troubleshooting and log analysis. Main use case of `Flow-ID` is to track service calls of our SaaS fashion commerce platform and initiated internal processing flows (executed synchronously via APIs or asynchronously via published events).

Data Definition

The `Flow-ID` must be passed through:

- RESTful API requests via `X-Flow-ID` proprietary header (see **MUST** propagate proprietary headers)
- Published events via `flow_id` event field (see metadata)

The following formats are allowed:

- `UUID` (RFC-4122)
- `base64` (RFC-4648)
- `base64url` (RFC-4648 Section 5)
- Random unique string restricted to the character set `[a-zA-Z0-9/+_-=]` maximal of 128 characters.

Note: If a legacy subsystem can only process `Flow-IDs` with a specific format or length, it must define this restriction in its API specification, and be generous and remove invalid characters or cut the length to the supported limit.

Hint: In case distributed tracing is supported by OpenTracing (internal link) you should ensure that created *spans* are tagged using `flow_id` — see How to Connect Log Output with OpenTracing Using Flow-IDs (internal link) or Best practices (internal link).

Service Guidance

- Services **must** support `Flow-ID` as generic input, i.e.

- RESTful API endpoints **must** support `X-Flow-ID` header in requests
- Event listeners **must** support the metadata `flow-id` from events.

Note: API-Clients **must** provide `Flow-ID` when calling a service or producing events. If no `Flow-ID` is provided in a request or event, the service must create a new `Flow-ID` .

- Services **must** propagate `Flow-ID` , i.e. use `Flow-ID` received with API calls or consumed events as…
  - input for all API called and events published during processing
  - data field written for logging and tracing

Hint: This rule also applies to application internal interfaces and events not published via Nakadi (but e.g. via AWS SQS, Kinesis or service specific DB solutions).

---

12. REST Design - Hypermedia

MUST use REST maturity level 2 [162]

We strive for a good implementation of REST Maturity Level 2 as it enables us to build resource-oriented APIs that make full use of HTTP verbs and status codes. You can see this expressed by many rules throughout these guidelines, e.g.:

- **MUST** avoid actions — think about resources
- **MUST** keep URLs verb-free
- **MUST** use HTTP methods correctly
- **SHOULD** only use most common HTTP status codes

Although this is not HATEOAS, it should not prevent you from designing proper link relationships in your APIs as stated in rules below.

MAY use REST maturity level 3 - HATEOAS [163]

We do not generally recommend to implement REST Maturity Level 3. HATEOAS comes with additional API complexity without real value in our SOA context where client and server interact via REST APIs and provide complex business functions as part of our e-commerce SaaS

platform.

Our major concerns regarding the promised advantages of HATEOAS (see also RESTistential Crisis over Hypermedia APIs, Why I Hate HATEOAS and others for a detailed discussion):

- We follow the API First principle with APIs explicitly defined outside the code with standard specification language. HATEOAS does not really add value for SOA client engineers in terms of API self-descriptiveness: a client engineer finds necessary links and usage description (depending on resource state) in the API reference definition anyway.
- Generic HATEOAS clients which need no prior knowledge about APIs and explore API capabilities based on hypermedia information provided, is a theoretical concept that we haven't seen working in practice and does not fit to our SOA set-up. The OpenAPI description format (and tooling based on OpenAPI) doesn't provide sufficient support for HATEOAS either.
- In practice relevant HATEOAS approximations (e.g. following specifications like HAL or JSON API) support API navigation by abstracting from URL endpoint and HTTP method aspects via link types. So, Hypermedia does not prevent clients from required manual changes when domain model changes over time.
- Hypermedia make sense for humans, less for SOA machine clients. We would expect use cases where it may provide value more likely in the frontend and human facing service domain.
- Hypermedia does not prevent API clients to implement shortcuts and directly target resources without 'discovering' them.

However, we do not forbid HATEOAS; you could use it, if you checked its limitations and still see clear value for your usage scenario that justifies its additional complexity. If you use HATEOAS please share experience and present your findings in the API Guild (internal_link).

MUST use common hypertext controls [164]

When embedding links to other resources into representations you must use the common hypertext control object. It contains at least one attribute:

- `href` : The URI of the resource the hypertext control is linking to. All our API are using HTTP(s) as URI scheme.

In API that contain any hypertext controls, the attribute name `href` is reserved for usage within

hypertext controls.

The schema for hypertext controls can be derived from this model:

```yaml
HttpLink:
  description: A base type of objects representing links to resources.
  type: object
  properties:
    href:
      description: Any URI that is using http or https protocol
      type: string
      format: uri
  required:
    - href
```

The name of an attribute holding such a **`HttpLink`** object specifies the relation between the object that contains the link and the linked resource. Implementations should use names from the IANA Link Relation Registry whenever appropriate. As IANA link relation names use hyphen-case notation, while this guide enforces snake_case notation for attribute names, hyphens in IANA names have to be replaced with underscores (e.g. the IANA link relation type **`version-history`** would become the attribute **`version_history`**)

Specific link objects may extend the basic link type with additional attributes, to give additional information related to the linked resource or the relationship between the source resource and the linked one.

E.g. a service providing "Person" resources could model a person who is married with some other person with a hypertext control that contains attributes which describe the other person (**`id`** , **`name`**) but also the relationship "spouse" between the two persons (**`since`**):

```json
{
  "id": "446f9876-e89b-12d3-a456-426655440000",
  "name": "Peter Mustermann",
  "spouse": {
    "href": "https://...",
    "since": "1996-12-19",
    "id": "123e4567-e89b-12d3-a456-426655440000",
    "name": "Linda Mustermann"
  }
}
```

Hypertext controls are allowed anywhere within a JSON model. While this specification would allow HAL, we actually don't recommend/enforce the usage of HAL anymore as the structural

separation of meta-data and data creates more harm than value to the understandability and usability of an API.

SHOULD use simple hypertext controls for pagination and self-references [165]

For pagination and self-references a simplified form of the extensible common hypertext controls should be used to reduce the specification and cognitive overhead. It consists of a simple URI value in combination with the corresponding link relations, e.g. `next`, `prev`, `first`, `last`, or `self`.

See **MUST** use common hypertext controls and **SHOULD** use pagination links for more information and examples.

MUST use full, absolute URI for resource identification [217]

Links to other resource must always use full, absolute URI.

Motivation: Exposing any form of relative URI (no matter if the relative URI uses an absolute or relative path) introduces avoidable client side complexity. It also requires clarity on the base URI, which might not be given when using features like embedding subresources. The primary advantage of non-absolute URI is reduction of the payload size, which is better achievable by following the recommendation to use gzip compression

MUST not use link headers with JSON entities [166]

For flexibility and precision, we prefer links to be directly embedded in the JSON payload instead of being attached using the uncommon link header syntax. As a result, the use of the `Link` Header defined by RFC 8288 in conjunction with JSON media types is forbidden.

---

13. REST Design - Performance

SHOULD reduce bandwidth needs and improve responsiveness [155]

APIs should support techniques for reducing bandwidth based on client needs. This holds for APIs that (might) have high payloads and/or are used in high-traffic scenarios like the public

Internet and telecommunication networks. Typical examples are APIs used by mobile web app clients with (often) less bandwidth connectivity. (Zalando is a 'Mobile First' company, so be mindful of this point.)

Common techniques include:

- compression of request and response bodies (see **SHOULD** use `gzip` compression)
- querying field filters to retrieve a subset of resource attributes (see **SHOULD** support partial responses via filtering below)
- `ETag` and `If-Match` / `If-None-Match` headers to avoid re-fetching of unchanged resources (see **MAY** consider to support `ETag` together with `If-Match` / `If-None-Match` header)
- `Prefer` header with `return=minimal` or `respond-async` to anticipate reduced processing requirements of clients (see **MAY** consider to support `Prefer` header to handle processing preferences)
- REST Design - Pagination for incremental access of larger collections of data items
- caching of master data items, i.e. resources that change rarely or not at all after creation (see **MUST** document cacheable `GET`, `HEAD`, and `POST` endpoints).

Each of these items is described in greater detail below.

SHOULD use `gzip` compression [156]

A servers and clients should support `gzip` content encoding to reduce the data transported over the network and thereby speed up response times, unless there is a good reason against this. Good reasons against compression are:

1. The content is already compressed, or
2. The server has not enough resources to support compression.

While `gzip` content encoding should be the default, servers must also support unencoded content for testing. This is ensured by content negotiation using the `Accept-Encoding` request header (see RFC 9110 Section 12.5.3). Successful compression is signaled via the `Content-Encoding` response header (see RFC 9110 Section 8.4). Clients and servers may support other compression algorithms as `compress`, `deflate`, or `br`.

To signal server support for compression the API specification should define both, the `Accept-`

`Encoding` request header and the **`Content-Encoding`** response header. This can be done by simply referencing the standard header definitions as follows (see also the default definition below):

```yaml
components:
  parameters|headers:
    Accept-Encoding:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/Accept-Encoding'
    Content-Encoding:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/Content-Encoding'
```

```yaml
components:
  parameters:
    Accept-Encoding:
        name: Accept-Encoding
        in: header
        required: false
        description: |-
          The **Accept-Encoding** indicates the content encoding (usually a
          compression algorithm) the client understands. The server selects one of
          the proposed encodings and returns his choice in the **Content-Encoding**
          response header.

          Supported compression algorithms of the server are [**gzip**][gzip] and
          **identity**, but other encodings from list below may be support too:

          * **gzip** - a compression format that uses [Lempel-Ziv coding][gzip]
            (LZ77) with a 32-bit CRC.
          * **compress** - a format using the [Lempel-Ziv-Welch][lzw] (LZW)
            algorithm.
          * **deflate** - A compression format that uses the [zlib][zlib] structure
            with the [deflate][deflate] compression algorithm.
          * **br** - a compression format that uses the [Brotli][brotli] algorithm.
          * **identity** - the identity function without modification or compression
            used to indicate unchanged responses. This value is always considered as
            acceptable, even if omitted.

          The server may choose not to compress the body of the response, if the
          identity value is acceptable, e.g. if the content is already compressed
          (e.g. for JPEG content), or if the server is missing resources to perform
          the compression operation.

          **Note:** A client is allowed to explicitly forbid the identity value by
          setting **identity;q=0** or **\*;q=0**, however, the server in return may
          respond with a **406** (Not Acceptable) error.

          See [RFC 9110 Section 12.5.3][rfc-9110-12.5.3] as well as [API Guideline
          Rule #156][api-156] for further details.

          [rfc-9110-12.5.3]: <https://tools.ietf.org/html/rfc9110#section-12.5.3>
```

```
          [api-156]: <https://opensource.zalando.com/restful-api-guidelines/#156>
          [gzip]: <https://en.wikipedia.org/wiki/LZ77_and_LZ78#LZ77>
          [lzw]: <https://en.wikipedia.org/wiki/LZW>
          [zlib]: <https://en.wikipedia.org/wiki/Zlib>
          [deflate]: <https://en.wikipedia.org/wiki/DEFLATE>
          [brotli]: <https://en.wikipedia.org/wiki/Brotli>
        schema:
          type: array
          items:
            type: string
        style: simple
        explode: false
        example: [ "gzip", "gzip;q=0.8,deflate;q=0.2", "gzip;q=0.9,identity;q=0" ]

components:
  headers:
    Content-Encoding:
        required: false
        description: |-
          The **Content-Encoding** header lists any encoding (usually a compression
          algorithm) applied to the message payload - in execution order. This allows
          the recipient to decode the content in order to obtain the original payload
          format.

          Supported compression algorithms of the server are [**gzip**][gzip] and
          **identity**, but other encoding from list below may be support too:

          * **gzip** - a compression format that uses [Lempel-Ziv coding][gzip]
            (LZ77) with a 32-bit CRC.
          * **compress** - a format using the [Lempel-Ziv-Welch][lzw] (LZW)
            algorithm.
          * **deflate** - A compression format that uses the [zlib][zlib] structure
            with the [deflate][deflate] compression algorithm.
          * **br** - a compression format that uses the [Brotli][brotli] algorithm.
          * **identity** - the identity function without modification or compression
            used to indicate unchanged responses. This value is always considered as
            acceptable, even if omitted.

          **Note:** the original media/content type is specified in the
          **Content-Type**-header, while the **Content-Encoding** applies to the
          representation of the data. If the original media is encoded in some way,
          e.g. as a zip file, then this information is not be included in the
          **Content-Encoding** header.

          See [RFC 9110 Section 8.4][rfc-9110-8.4] as well as [API Guideline Rule
          #156][api-156] for further details.

          [rfc-9110-8.4]: <https://tools.ietf.org/html/rfc9110#section-8.4>
          [api-156]: <https://opensource.zalando.com/restful-api-guidelines/#156>
          [gzip]: <https://en.wikipedia.org/wiki/LZ77_and_LZ78#LZ77>
          [lzw]: <https://en.wikipedia.org/wiki/LZW>
          [zlib]: <https://en.wikipedia.org/wiki/Zlib>
```

```
          [deflate]: <https://en.wikipedia.org/wiki/DEFLATE>
          [brotli]: <https://en.wikipedia.org/wiki/Brotli>
        schema:
          type: array
          items:
            type: string
        style: simple
        explode: false
        example: [ "gzip", "deflate", "gzip,deflate" ]
```

Note: since most server and client frameworks still **do not** support `gzip` -compression **out-of-the-box** you need to manually activate it, e.g. Spring Boot, or add a dedicated middleware, e.g. gin-gionic/gin.

SHOULD support partial responses via filtering [157]

Depending on your use case and payload size, you can significantly reduce network bandwidth need by supporting filtering of returned entity fields. Here, the client can explicitly determine the subset of fields he wants to receive via the `fields` query parameter. (It is analogue to GraphQL `fields` and simple queries, and also applied, for instance, for Google Cloud API's partial responses.)

Unfiltered

```
GET http://api.example.org/users/123 HTTP/1.1

HTTP/1.1 200 OK
Content-Type: application/json

{
  "id": "cddd5e44-dae0-11e5-8c01-63ed66ab2da5",
  "name": "John Doe",
  "address": "1600 Pennsylvania Avenue Northwest, Washington, DC, United States",
  "birthday": "1984-09-13",
  "friends": [ {
    "id": "1fb43648-dae1-11e5-aa01-1fbc3abb1cd0",
    "name": "Jane Doe",
    "address": "1600 Pennsylvania Avenue Northwest, Washington, DC, United States",
    "birthday": "1988-04-07"
  } ]
}
```

Filtered

```
GET http://api.example.org/users/123?fields=(name,friends(name)) HTTP/1.1

HTTP/1.1 200 OK
```

```
Content-Type: application/json

{
  "name": "John Doe",
  "friends": [ {
    "name": "Jane Doe"
  } ]
}
```

The `fields` query parameter determines the fields returned with the response payload object. For instance, `(name)` returns `users` root object with only the `name` field, and `(name,friends(name))` returns the `name` and the nested `friends` object with only its `name` field.

OpenAPI doesn't support you in formally specifying different return object schemes depending on a parameter. When you define the field parameter, we recommend to provide the following description: **`Endpoint supports filtering of return object fields as described in [Rule #157](https://opensource.zalando.com/restful-api-guidelines/#157)`**

The syntax of the query `fields` value is defined by the following BNF grammar.

```
<fields>            ::= [ <negation> ] <fields_struct>
<fields_struct>     ::= "(" <field_items> ")"
<field_items>       ::= <field> [ "," <field_items> ]
<field>             ::= <field_name> | <fields_substruct>
<fields_substruct>  ::= <field_name> <fields_struct>
<field_name>        ::= <dash_letter_digit> [ <field_name> ]
<dash_letter_digit> ::= <dash> | <letter> | <digit>
<dash>              ::= "-" | "_"
<letter>            ::= "A" | ... | "Z" | "a" | ... | "z"
<digit>             ::= "0" | ... | "9"
<negation>          ::= "!"
```

Note: Following the principle of least astonishment, you should not define the `fields` query parameter using a default value, as the result is counter-intuitive and very likely not anticipated by the consumer.

SHOULD allow optional embedding of sub-resources [158]

Embedding related resources (also know as *Resource expansion*) is a great way to reduce the number of requests. In cases where clients know upfront that they need some related resources they can instruct the server to prefetch that data eagerly. Whether this is optimized on the server, e.g. a database join, or done in a generic way, e.g. an HTTP proxy that transparently embeds

resources, is up to the implementation.

See **MUST** stick to conventional query parameters for naming, e.g. "embed" for steering of embedded resource expansion. Please use the BNF grammar, as already defined above for filtering, when it comes to an embedding query syntax.

Embedding a sub-resource can possibly look like this where an order resource has its order items as sub-resource (/order/{orderId}/items):

```
GET /order/123?embed=(items) HTTP/1.1

{
  "id": "123",
  "_embedded": {
    "items": [
      {
        "position": 1,
        "sku": "1234-ABCD-7890",
        "price": {
          "amount": 71.99,
          "currency": "EUR"
        }
      }
    ]
  }
}
```

MUST document cacheable `GET`, `HEAD`, and `POST` endpoints [227]

Caching has to take many aspects into account, e.g. general cacheability of response information, our guideline to protect endpoints using SSL and OAuth authorization, resource update and invalidation rules, existence of multiple consumer instances. Caching is in best case complex, e.g. with respect to consistency, in worst case inefficient.

As a consequence, client side as well as transparent web caching should be avoided, unless the service supports and requires it to protect itself, e.g. in case of a heavily used and therefore rate limited master data service, i.e. data items that rarely or not at all change after creation.

As default, servers and clients should always set the `Cache-Control` header to `Cache-Control: no-store` and assume the same setting, if no `Cache-Control` header is provided.

Note: There is no need to document this default setting. However, please make sure that your

framework is attaching these header values by default, or ensure this manually, e.g. using the best practice of Spring Security as shown below. Any setup deviating from this default must be sufficiently documented.

```
Cache-Control: no-cache, no-store, must-revalidate, max-age=0
```

If your service really requires to support caching, please observe the following rules:

- Document all cacheable `GET`, `HEAD`, and `POST` endpoints by declaring the support of `Cache-Control`, `Vary`, and `ETag` headers in response. **Note:** you must not define the `Expires` header to prevent redundant and ambiguous definition of cache lifetime. A sensible default documentation of these headers is given below.
- Take care to specify the ability to support caching by defining the right caching boundaries, i.e. time-to-live and cache constraints, by providing sensible values for `Cache-Control` and `Vary` in your service. We will explain best practices below.
- Provide efficient methods to warm up and update caches (see Cache Support Patterns).

Usually, you can reuse the standard `Cache-Control`, `Vary`, and `ETag` response header definitions provided by the guideline as follows:

```
components:
  parameters|headers:
    Cache-Control:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/Cache-Control'
    Vary:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/Vary'
    ETag:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/ETag'
```

See also **MAY** consider to support `ETag` together with `If-Match` / `If-None-Match` header.

Cache Support Patterns

To make best use of caching in micro service environments you need to provide efficient methods to warm up and update caches, e.g. as follows:

- In general, you should support `ETag` Together With `If-Match` / `If-None-Match` Header on all cacheable endpoints.
- For larger data items you should support `HEAD` requests or more a bit more efficient `GET` requests with `If-None-Match` header to check for updates.

- For small data sets you should provide a **get full collection** `GET` endpoint that supports an `ETag` for the collection in combination with a `HEAD` or `GET` requests with `If-None-Match` to check for updates.
- For medium sized data sets provide a **get full collection** `GET` endpoint that supports an `ETag` for the collection in combination with REST Design - Pagination and `<entity-tag>` filtering `GET` requests for limiting the response to changes since the provided `<entity-tag>`. **Note:** this is not supported by generic client and proxy caches on HTTP layer.

Hint: For proper cache support, you must return **304** without content on a failed `HEAD` or `GET` request with `If-None-Match: <entity-tag>` instead of **412**. For `ETag` see also **MAY** consider to support `ETag` together with `If-Match` / `If-None-Match` header.

Default Header Values

The default setting for `Cache-Control` should contain the `private` directive for endpoints with standard OAuth authorization, as well as the `must-revalidate` directive to ensure, that the client does not use stale cache entries. Last, the `max-age` directive should be set to a value between a few seconds (`max-age=60`) and a few hours (`max-age=86400`) depending on the change rate of your master data and your requirements to keep clients consistent.

```
Cache-Control: private, must-revalidate, max-age=300
```

The default setting for `Vary` is harder to determine correctly. It depends on the API endpoint, e.g. whether it supports compression, accepts different media types, or requires other request specific headers. To support correct caching you have to carefully choose the value. However, a good first default may be:

```
Vary: accept, accept-encoding
```

Anyhow, this is only relevant, if you encourage clients to install generic HTTP layer client and proxy caches.

Caching strategy

Generic client and proxy caching on HTTP level is hard to configure. Therefore, we strongly recommend to attach the (possibly distributed) cache directly to the service (or gateway) layer of your application. This relieves the service from interpreting the `Vary` header and greatly simplifies the usage patterns of the `Cache-Control` and `ETag` headers. Moreover, is highly efficient with respect to cache performance and overhead, and allows to support more advanced

cache update and warm up patterns.

Anyhow, please carefully read RFC 9111 before adding any client or proxy cache.

---

14. REST Design - Pagination

MUST support pagination [159]

Access to lists of data items must support pagination to protect the service against overload as well as to support client side iteration and batch processing experience. This holds true for all lists that are (potentially) larger than just a few hundred entries.

There are two well known page iteration techniques:

- **Offset-based pagination**: numeric offset identifies the first page-entry
- **Cursor-based pagination** — aka key-based pagination: a unique key identifies the first page-entry (see also Twitter API or Facebook API)

The technical conception of pagination should also consider user experience (see Pagination Usability Findings In eCommerce), for instance, jumping to a specific page is far less used than navigation via `next` / `prev` page links (see **SHOULD** use pagination links). This favors an API design using cursor-based instead of offset-based pagination — see **SHOULD** prefer cursor-based pagination, avoid offset-based pagination.

Note: To provide a consistent look and feel of pagination patterns, you must stick to the common query parameter names defined in **MUST** stick to conventional query parameters.

SHOULD prefer cursor-based pagination, avoid offset-based pagination [160]

Cursor-based pagination is usually better and more efficient when compared to offset-based pagination. Especially when it comes to high-data volumes and/or storage in NoSQL databases.

Before choosing cursor-based pagination, consider the following trade-offs:

- Usability/framework support:
  - Offset-based pagination is more widely known than cursor-based pagination, so it has

more framework support and is easier to use for API clients.

- Use case - jump to a certain page:
  - If jumping to a particular page in a range (e.g., 51 of 100) is really a required use case, cursor-based navigation may not be feasible.
- Data changes may lead to anomalies in result pages:
  - Offset-based pagination may create duplicates or lead to missing entries if rows are inserted or deleted between two subsequent paging requests.
  - If implemented incorrectly, cursor-based pagination may fail when the cursor entry has been deleted before fetching the pages.
- Performance considerations - efficient server-side processing using offset-based pagination is hardly feasible for:
  - Very big data sets, especially if they cannot reside in the main memory of the database.
  - Sharded or NoSQL databases.

The `cursor` used for pagination is an opaque pointer to a page, that must never be **inspected** or **constructed** by clients. It usually encodes (encrypts) the page position, i.e. the unique identifier of the first or last page element, the pagination direction, and the applied query filters (or a hash over these) to safely recreate the collection (see also best practice Cursor-based pagination in RESTful APIs below).

SHOULD use pagination response page object [248]

For iterating over collections (result sets) we propose to either use cursors (see **SHOULD** prefer cursor-based pagination, avoid offset-based pagination) or simple hypertext control links (see **SHOULD** use pagination links). To implement these in a consistent way, we have defined a response page object pattern with the following field semantics:

- `self` : the link or cursor pointing to the same page.
- `first` : the link or cursor pointing to the first page.
- `prev` : the link or cursor pointing to the previous page. It is not provided, if it is the first page.
- `next` : the link or cursor pointing to the next page. It is not provided, if it is the last page.
- `last` : the link or cursor pointing to the last page.

Pagination responses should contain the following additional array field to transport the page content:

- **`items`** : array of resources, holding all the items of the current page (**`items`** may be replaced by a resource name).

For responses to **`GET with body`** operations, the applied query filters **should** be (and for normal **`GET`** may be) returned using the following field:

- **`query`** : object containing the query filters applied in the search request to filter the collection resource. This can be directly used as the request body when following the pagination links.

In conclusion, the standard response page using plain cursors or pagination links may be defined as follows:

```
ResponsePage:
  type: object
  required:
    - items
  properties:
    self:
      description: Pagination link|cursor pointing to the current page.
      type: string
      format: uri|cursor
    first:
      description: Pagination link|cursor pointing to the first page.
      type: string
      format: uri|cursor
    prev:
      description: Pagination link|cursor pointing to the previous page.
      type: string
      format: uri|cursor
    next:
      description: Pagination link|cursor pointing to the next page.
      type: string
      format: uri|cursor
    last:
      description: Pagination link|cursor pointing to the last page.
      type: string
      format: uri|cursor

    query:
      description: >
        Object containing the query filters applied to the collection resource.
        This can be directly used as a request body (together with the cursors/links)
        when requesting other pages.
      type: object
```

```
      properties: ...

    items:
      description: Array of collection items.
      type: array
      required: false
      items:
        type: ...
```

Note: While you may support plain cursors for `next`, `prev`, `first`, `last`, and `self`, it is best practice to replace these with pagination links — see **SHOULD** use pagination links.

SHOULD use pagination links [161]

To simplify client design, APIs should support simplified hypertext controls as standard pagination links where applicable:

```
{
  "self": "https://my-service.zalandoapis.com/resources?cursor=<self-position>",
  "first": "https://my-service.zalandoapis.com/resources?cursor=<first-position>",
  "prev": "https://my-service.zalandoapis.com/resources?cursor=<previous-position>",
  "next": "https://my-service.zalandoapis.com/resources?cursor=<next-position>",
  "last": "https://my-service.zalandoapis.com/resources?cursor=<last-position>",
  "query": {
    "query-param-<1>": ...,
    "query-param-<n>": ...
  },
  "items": [...]
}
```

For `GET with body` operations, the `query` object can to be used as a request body with either of these links. (It should be equivalent to just resend the original body.)

See also **SHOULD** use pagination response page object for details on the pagination fields and page result object.

SHOULD avoid a total result count [254]

In pagination responses you should generally avoid providing a *total result count*, since calculating it is a costly operation that is usually not required by clients. Counting the total number of results for complex queries usually requires a full scan of all involved indexes, as it is difficult to calculate and cache it in advance. While this is only an implementation detail, it is important to consider that providing these total counts over the life-span of a service might become expensive as the data set grows over time.

As clients may integrate against these counts over time alongside data set growth, removing them will be more difficult than not providing them in the first place.

If your consumer really requires a total result count in the response, you may support this requirement via the `Prefer` header adding the directive `return=total-count` (see also **MAY** consider to support `Prefer` header to handle processing preferences).

---

15. REST Design - Compatibility

MUST not break backward compatibility [106]

Change APIs, but keep all consumers running. Consumers usually have independent release lifecycles, focus on stability, and avoid changes that do not provide additional value. APIs are contracts between service providers and service consumers that cannot be broken via unilateral decisions.

There are two techniques to change APIs without breaking them:

- follow rules for compatible extensions
- introduce new API versions and still support older versions with deprecation

We strongly encourage using compatible API extensions and discourage versioning (see **SHOULD** avoid versioning and **MUST** use media type versioning below). The following guidelines for service providers (**SHOULD** prefer compatible extensions) and consumers (**MUST** prepare clients to accept compatible API extensions) enable us (having Postel's Law in mind) to make compatible changes without versioning.

Note: There is a difference between incompatible and breaking changes. Incompatible changes are changes that are not covered by the compatibility rules below. Breaking changes are incompatible changes deployed into operation, and thereby breaking running API consumers. Usually, incompatible changes are breaking changes when deployed into operation. However, in specific controlled situations it is possible to deploy incompatible changes in a non-breaking way, if no API consumer is using the affected API aspects (see also Deprecation guidelines).

Hint: Please note that the compatibility guarantees are for the "on the wire" format. Binary or source compatibility of code generated from an API definition is not covered by these rules. If client implementations update their generation process to a new version of the API definition, it

has to be expected that code changes are necessary.

SHOULD prefer compatible extensions [107]

API designers should apply the following rules to evolve RESTful APIs for services in a backward-compatible way:

- Add only optional, never mandatory fields.
- Never change the semantic of fields (e.g. changing the semantic from customer-number to customer-id, as both are different unique customer keys)
- Input fields may have (complex) constraints being validated via server-side business logic. Never change the validation logic to be more restrictive and make sure that all constraints are clearly defined in description.
- `enum` ranges can be reduced when used as input parameters, only if the server is ready to accept and handle old range values too. The range can be reduced when used as output parameters.
- `enum` ranges cannot be extended when used for output parameters — clients may not be prepared to handle it. However, enum ranges can be extended when used for input parameters.
- You **SHOULD** use open-ended list of values (`x-extensible-enum`) for enumeration types that are used for output parameters and likely to be extended with growing functionality. The API definition should be updated first before returning new values.
- Consider **SHOULD** be aware of services not fully supporting JSON/unicode in case a URL has to change.

SHOULD design APIs conservatively [109]

Designers of service provider APIs should be conservative and accurate in what they accept from clients:

- Unknown input fields in payload or URL should not be ignored; servers should provide error feedback to clients via an HTTP 400 response code. Otherwise, if unexpected fields are planned to be handled in some way instead of being rejected, API designers **must** document clearly how unexpected fields are being handled for `PUT` , `POST` , and `PATCH` requests.
- Be accurate in defining input data constraints (like formats, ranges, lengths etc.) — and check constraints and return dedicated error information in case of violations.

- Prefer being more specific and restrictive (if compliant to functional requirements), e.g. by defining length range of strings. It may simplify implementation while providing freedom for further evolution as compatible extensions.

Not ignoring unknown input fields is a specific deviation from Postel's Law (e.g. see also The Robustness Principle Reconsidered) and a strong recommendation. Servers might want to take different approach but should be aware of the following problems and be explicit in what is supported:

- Ignoring unknown input fields is actually not an option for `PUT`, since it becomes asymmetric with subsequent `GET` response and HTTP is clear about the `PUT` *replace* semantics and default roundtrip expectations (see RFC 9110 Section 9.3.4). Note, accepting (i.e. not ignoring) unknown input fields and returning it in subsequent `GET` responses is a different situation and compliant to `PUT` semantics.
- Certain client errors cannot be recognized by servers, e.g. attribute name typing errors will be ignored without server error feedback. The server cannot differentiate between the client intentionally providing an additional field versus the client sending a mistakenly named field, when the client's actual intent was to provide an optional input field.
- Future extensions of the input data structure might be in conflict with already ignored fields and, hence, will not be compatible, i.e. break clients that already use this field but with different type.

In specific situations, where a (known) input field is not needed anymore, it either can stay in the API definition with "not used anymore" description or can be removed from the API definition as long as the server ignores this specific parameter.

MUST prepare clients to accept compatible API extensions [108]

Service clients should apply the robustness principle:

- Be conservative with API requests and data passed as input, e.g. avoid to exploit definition deficits like passing megabytes of strings with unspecified maximum length.
- Be tolerant in processing and reading data of API responses, more specifically service clients must be prepared for compatible API extensions of response data:
  - Be tolerant with unknown fields in the payload (see also Fowler's "TolerantReader" post), i.e. ignore new fields but do not eliminate them from payload if needed for subsequent `PUT` requests.

- Be prepared that `x-extensible-enum` return parameters (see rule 112) may deliver new values; either be agnostic or provide default behavior for unknown values, and do not eliminate them.
- Be prepared to handle HTTP status codes not explicitly specified in endpoint definitions. Note also, that status codes are extensible. Default handling is how you would treat the corresponding **2xx** code (see RFC 9110 Section 15).
- Follow the redirect when the server returns HTTP status code **301** (Moved Permanently).

The Handling compatible API extensions section describes a best practice to implement the requirements in Java.

MUST treat OpenAPI specification as open for extension by default [111]

The OpenAPI specification is not very specific on default extensibility of objects, and redefines JSON-Schema keywords related to extensibility, like `additionalProperties`. Following our compatibility guidelines, OpenAPI object definitions are considered open for extension by default as per Section 5.18 "additionalProperties" of JSON-Schema.

When it comes to OpenAPI, this means an `additionalProperties` declaration is not required to make an object definition extensible:

- API clients consuming data must not assume that objects are closed for extension in the absence of an `additionalProperties` declaration and must ignore fields sent by the server they cannot process. This allows API servers to evolve their data formats.
- For API servers receiving unexpected data, the situation is slightly different. According to **SHOULD** design APIs conservatively, instead of ignoring fields, servers *should* reject requests whose entities contain undefined fields in order to signal to clients that those fields would not be stored on behalf of the client. Otherwise, if unexpected fields are planned to be handled in some way instead of being rejected, API designers **must** document clearly how unexpected fields are being handled for `PUT`, `POST`, and `PATCH` requests.

API formats must not declare `additionalProperties` to be false, as this prevents objects being extended in the future.

Note that this guideline concentrates on default extensibility and does not exclude the use of `additionalProperties` with a schema as a value, which might be appropriate in some

circumstances, e.g. see **SHOULD** define maps using `additionalProperties`.

SHOULD avoid versioning [113]

When changing your RESTful APIs, do so in a compatible way and avoid generating additional API versions. Multiple versions can significantly complicate understanding, testing, maintaining, evolving, operating and releasing our systems (supplementary reading).

If changing an API can't be done in a compatible way, then proceed in one of these three ways:

- create a new resource (variant) in addition to the old resource variant
- create a new service endpoint — i.e. a new application with a new API (with a new domain name)
- create a new API version supported in parallel with the old API by the same microservice

As we discourage versioning by all means because of the manifold disadvantages, we strongly recommend to only use the first two approaches.

MUST use media type versioning [114]

However, when API versioning is unavoidable, you have to design your multi-version RESTful APIs using media type versioning (see **MUST** not use URL versioning). Media type versioning is less tightly coupled since it supports content negotiation and hence reduces complexity of release management.

Version information and media type are provided together via the HTTP `Content-Type` header — e.g. `application/x.zalando.cart+json;version=2`. For incompatible changes, a new media type version for the resource is created. To generate the new representation version, consumer and producer can do content negotiation using the HTTP `Content-Type` and `Accept` headers.

This versioning method only applies to the request and response payload schema, not to URI or method semantics.

Custom media type format

Custom media type format should have the following pattern:

```
application/x.<custom-media-type>+json;version=<version>
```

- `<custom-media-type>` is a custom type name, e.g. `x.zalando.cart`
- `<version>` is a (sequence) number, e.g. `2`

Example

In this example, a client wants only the new version of the response:

```
Accept: application/x.zalando.cart+json;version=2
```

A server responding to this, as well as a client sending a request with content should use the `Content-Type` header, declaring that one is sending the new version:

```
Content-Type: application/x.zalando.cart+json;version=2
```

Media type versioning should…

- Use a custom media type, e.g. `application/x.zalando.cart+json`
- Include media type versions in request and response headers to increase visibility
- Include `Content-Type` in the `Vary` header to enable proxy caches to differ between versions

```
Vary: Content-Type
```

Until an incompatible change is necessary, it is recommended to stay with the standard `application/json` media type without versioning.

Further reading: API Versioning Has No "Right Way" provides an overview on different versioning approaches to handle breaking changes without being opinionated.

MUST not use URL versioning [115]

With URL versioning a (major) version number is included in the path, e.g. `/v1/customers` . The consumer has to wait until the provider has been released and deployed. If the consumer also supports hypermedia links — even in their APIs — to drive workflows (HATEOAS), this quickly becomes complex. So does coordinating version upgrades — especially with hyperlinked service dependencies — when using URL versioning. To avoid this tighter coupling and complexer release management we do not use URL versioning, instead we **MUST** use media type versioning with content negotiation.

MUST always return JSON objects as top-level data structures [110]

In a response body, you must always return a JSON object (and not e.g. an array) as a top level data structure to support future extensibility. JSON objects support compatible extension by additional attributes. This allows you to easily extend your response and e.g. add pagination later, without breaking backwards compatibility. See **SHOULD** use pagination links for an example.

Maps (see **SHOULD** define maps using `additionalProperties`), even though technically objects, are also forbidden as top level data structures, since they don't support compatible, future extensions.

SHOULD use open-ended list of values (`x-extensible-enum`) for enumeration types [112]

JSON schema `enum` is per definition a closed set of values that is assumed to be complete and not intended for extension. This closed principle of enumerations imposes compatibility issues when an enumeration must be extended. To avoid these issues, we recommend to use an open-ended list of values instead of an enumeration unless:

1. the API has full control of the enumeration values, i.e. the list of values does not depend on any external tool or interface, and
2. the list of values is complete with respect to any thinkable and unthinkable future feature.

To specify an open-ended list of values via the `x-extensible-enum` property as follows:

```
delivery_methods:
  type: string
  x-extensible-enum:
    - PARCEL
    - LETTER
    - EMAIL
```

Note: `x-extensible-enum` is a proprietary extension of the JSON Schema standard that is e.g. visible via Swagger UI, but ignored by most other tools.

See **SHOULD** declare enum values using UPPER_SNAKE_CASE string about enum value naming conventions.

Note, `x-extensible-enum` is a different concept than JSON schema `examples` which is just a

list of a few example values, whereas `x-extensible-enum` defines all valid values (for a specific API and service version) and has the advantage of an extensible full type-range specification that is validated by the service.

Important: Clients must be prepared for extensions of enums returned with server responses, i.e. must implement a fallback / default behavior to handle unknown new enum values — see **MUST** prepare clients to accept compatible API extensions. API owners must take care to extend enums in a compatible way that does not change the semantics of already existing enum values, for instance, do not split an old enum value into two new enum values. Services should only extend `x-extensible-enum` ranges, and only accept and return values listed in the API definition, i.e. the API definition needs to be updated first before the service accepts/returns new values — see also **SHOULD** prefer compatible extensions.

---

16. REST Design - Deprecation

Sometimes it is necessary to phase out an API endpoint, an API version, or an API feature, e.g. if a field or parameter is no longer supported or a whole business functionality behind an endpoint is supposed to be shut down. As long as the API endpoints and features are still used by consumers these shut downs are breaking changes and not allowed. To progress the following deprecation rules have to be applied to make sure that the necessary consumer changes and actions are well communicated and aligned using *deprecation* and *sunset* dates.

MUST reflect deprecation in API specifications [187]

The API deprecation must be part of the API specification.

If an API endpoint (operation object), an input argument (parameter object), an in/out data object (schema object), or on a more fine grained level, a schema attribute or property should be deprecated, the producers must set `deprecated: true` for the affected element and add further explanation to the `description` section of the API specification. If a future shut down is planned, the producer must provide a sunset date and document in details what consumers should use instead and how to migrate.

MUST obtain approval of clients before API shut down [185]

Before shutting down an API, version of an API, or API feature the producer must make sure, that all clients have given their consent on a sunset date. Producers should help consumers to

migrate to a potential new API or API feature by providing a migration manual and clearly state the time line for replacement availability and sunset (see also **SHOULD** add `Deprecation` and `Sunset` header to responses). Once all clients of a sunset API feature are migrated, the producer may shut down the deprecated API feature.

MUST collect external partner consent on deprecation time span [186]

If the API is consumed by any external partner, the API owner must define a reasonable time span that the API will be maintained after the producer has announced deprecation. All external partners must state consent with this after-deprecation-life-span, i.e. the minimum time span between official deprecation and first possible sunset, **before** they are allowed to use the API.

MUST monitor usage of deprecated API scheduled for sunset [188]

Owners of an API, API version, or API feature used in production that is scheduled for sunset must monitor the usage of the sunset API, API version, or API feature in order to observe migration progress and avoid uncontrolled breaking effects on ongoing consumers. See also **SHOULD** monitor API usage.

SHOULD add `Deprecation` and `Sunset` header to responses [189]

During the deprecation phase, the producer should add a `Deprecation: <date-time>` (see draft: RFC Deprecation HTTP Header Field) and - if also planned - a `Sunset: <date-time>` (see RFC 8594) header on each response affected by a deprecated element (see **MUST** reflect deprecation in API specifications).

The `Deprecation` header can either be set to `true` - if a feature is retired -, or carry a deprecation time stamp, at which a replacement will become/became available and consumers must not on-board any longer (see **MUST** not start using deprecated APIs). The optional `Sunset` time stamp carries the information when consumers latest have to stop using a feature. The sunset date should always offer an eligible time interval for switching to a replacement feature.

```
Deprecation: Tue, 31 Dec 2024 23:59:59 GMT
Sunset: Wed, 31 Dec 2025 23:59:59 GMT
```

If multiple elements are deprecated the `Deprecation` and `Sunset` headers are expected to be

set to the earliest time stamp to reflect the shortest interval at which consumers are expected to get active. The **`Deprecation`** and **`Sunset`** headers can be defined as follows in the API specification (see also the default definition below):

```
components:
  parameters|headers:
    Deprecation:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/Deprecation'
    Sunset:
      $ref: 'https://opensource.zalando.com/restful-api-guidelines/models/headers-1.0.0.yaml#/Sunset'
```

```
components:
  headers:
    Deprecation:
        required: false
        description: |-
          The **Deprecation** response header (see [Draft "The Deprecation HTTP
          Header Field"][draft]) announces an upcoming deprecation of a feature or
          resource. The deprecation value is either a timestamp as defined in [RFC
          9110 Section 5.6.7][rfc-9110-5.6.7] or `true` - if the feature is already
          deprecated.

          The scope of the distinct deprecated feature or resource must be derived
          from the API spec. If the timestamp points to the past, the API spec also
          contains a migration advise.

          Clients should monitor the usage of **Deprecation** headers and notify
          about them in time, so that migration measures can be planned and executed
          timely (see also [API Guideline - Deprecation][api]).

          [draft]: <https://datatracker.ietf.org/doc/html/draft-ietf-httpapi-deprecation-header-02>
          [rfc-9110-5.6.7]: <https://tools.ietf.org/html/rfc9110#section-5.6.7>
          [api]: <https://opensource.zalando.com/restful-api-guidelines/#deprecation>
        schema:
          type: string
          format: date-time
        example: [ "Tue, 31 Dec 2024 23:59:59 GMT" ]

    Sunset:
        required: false
        description: |-
          The **Sunset** response header (see [RFC 8594][rfc-8594]) communicates the
          point in time at which the feature or resource becomes unresponsive. The
          sunset value is providing a timestamp as defined in [RFC 9110 Section
          5.6.7][rfc-9110-5.6.7] usually pointing to the future. If a sunset value
          points to the past, the feature or resource must be expected to become
          unavailable at any time.

          Clients should monitor the usage of **Sunset** headers and warn/alert about
          them before the sunset time has come to take counter measures, e.g. prepare
```

```
        a client shutdown or migration (see also [API Guideline -
        Deprecation][api-deprecation]).

        [rfc-8594]: <https://tools.ietf.org/html/rfc8594>
        [rfc-9110-5.6.7]: <https://tools.ietf.org/html/rfc9110#section-5.6.7>
        [api-deprecation]: <https://opensource.zalando.com/restful-api-guidelines/#deprecation>
      schema:
        type: string
        format: date-time
      example: "Wed, 31 Dec 2025 23:59:59 GMT"
```

Note: adding the `Deprecation` and `Sunset` header is not sufficient to gain client consent to shut down an API or feature.

Hint: In earlier guideline versions, we used the `Warning` header to provide the deprecation info to clients. However, `Warning` header has a less specific semantics, will be obsolete with draft: RFC HTTP Caching, and our syntax was not compliant with RFC 9111 Section 5.5 "Warning".

SHOULD add monitoring for `Deprecation` and `Sunset` header [190]

Clients should monitor the `Deprecation` and `Sunset` headers in HTTP responses to get information about future sunset of APIs and API features (see SHOULD add `Deprecation` and `Sunset` header to responses). We recommend that client owners build alerts on this monitoring information to ensure alignment with service owners on required migration task.

Hint: In earlier guideline versions, we used the `Warning` header to provide the deprecation info (see hint in SHOULD add `Deprecation` and `Sunset` header to responses).

MUST not start using deprecated APIs [191]

Clients must not start using deprecated APIs, API versions, or API features.

---

17. REST Operation

MUST publish OpenAPI specification for non-component-internal APIs [192]

All service applications must publish OpenAPI specifications of their external APIs. While this is optional for internal APIs, i.e. APIs marked with the **component-internal** API audience group, we

still recommend to do so to profit from the API management infrastructure.

An API is published by copying its **OpenAPI specification** into the reserved **/zalando-apis** directory of the **deployment artifact** used to deploy the provisioning service. The directory must only contain **self-contained YAML files** that each describe one API (exception see **MUST** only use durable and immutable remote references). We prefer this deployment artifact-based method over the past (now legacy) `.well-known/schema-discovery` service endpoint-based publishing process, that we only support for backward compatibility reasons.

Background: In our dynamic and complex service infrastructure, it is important to provide API client developers a central place with online access to the API specifications of all running applications. As a part of the infrastructure, the API publishing process is used to detect API specifications. The findings are published in the API Portal - the universal hub for all Zalando APIs.

Note: To publish an API, it is still necessary to deploy the artifact successful, as we focus the discovery experience on APIs supported by running services.

SHOULD monitor API usage [193]

Owners of APIs used in production should monitor API service to get information about its using clients. This information, for instance, is useful to identify potential review partner for API changes.

Hint: A preferred way of client detection implementation is by logging of the client-id retrieved from the OAuth token.

---

18. EVENT Basics - Event Types

Zalando's architecture centers around decoupled microservices and in that context we favour asynchronous event driven approaches. The guidelines focus on how to design and publish events intended to be shared for others to consume.

Events are defined using an item called an *Event Type*. The Event Type allows events to have their structure declared with a schema by producers and understood by consumers. An Event Type declares standard information, such as a name, an owning application (and by implication, an owning team), a schema defining the event's custom data, and a compatibility mode declaring

how the schema will be evolved. Event Types also allow the declaration of validation and enrichment strategies for events, along with supplemental information such as how events can be partitioned in an event stream.

Event Types belong to a well known *Event Category* (such as a data change category), which provides extra information that is common to that kind of event.

Event Types can be published and made available as API resources for teams to use, typically in an *Event Type Registry*. Each event published can then be validated against the overall structure of its event type and the schema for its custom data.

The basic model described above was originally developed in the Nakadi project, which acts as a reference implementation (see also Nakadi API (internal_link)) of the event type registry, and as a validating publish/subscribe broker for event producers and consumers.

MUST define events compliant with overall API guidelines [208]

Events must be consistent with other API data and the API Guidelines in general (as far as applicable).

Everything expressed in the Introduction to these Guidelines is applicable to event data interchange between services. This is because our events, just like our APIs, represent a commitment to express what our systems do and designing high-quality, useful events allows us to develop new and interesting products and services.

What distinguishes events from other kinds of data is the delivery style used, asynchronous publish-subscribe messaging. But there is no reason why they could not be made available using a REST API, for example via a search request or as a paginated feed, and it will be common to base events on the models created for the service's REST API.

The following existing guideline sections are applicable to events:

- General guidelines
- REST Basics - Data formats
- REST Basics - JSON payload
- REST Design - Hypermedia

MUST treat events as part of the service interface [194]

Events are part of a service's interface to the outside world equivalent in standing to a service's REST API. Services publishing data for integration must treat their events as a first class design concern, just as they would an API. For example this means approaching events with the "API first" principle in mind as described in the Introduction.

MUST make event schema available for review [195]

Services publishing event data for use by others must make the event schema as well as the event type definition available for review.

MUST specify and register events as event types [197]

In Zalando's architecture, events are registered using a structure called an *Event Type*. The Event Type declares standard information as follows:

- A well known event category, such as a general or data change category.
- The name of the event type.
- The definition of the event target audience.
- An owning application, and by implication, an owning team.
- A schema defining the event payload.
- The compatibility mode for the type.

Event Types allow easier discovery of event information and ensure that information is well-structured, consistent, and can be validated. The core Event Type structure is shown below as an OpenAPI object definition:

```
EventType:
  description: |
    An event type defines the schema and its runtime properties. The required
    fields are the minimum set the creator of an event type is expected to
    supply.
  required:
    - name
    - category
    - owning_application
    - schema
  properties:
    name:
      description: |
        Name of this EventType. The name must follow the functional naming
        pattern `<functional-name>.<event-name>` to preserve global
```

```
    uniqueness and readability.
  type: string
  pattern: '[a-z][a-z0-9-]*\.[a-z][a-z0-9-]*(\.[Vv][0-9.]+)?'
  example: |
    transactions-order.order-cancelled
    customer-personal-data.email-changed.v2
audience:
  type: string
  x-extensible-enum:
    - component-internal
    - business-unit-internal
    - company-internal
    - external-partner
    - external-public
  description: |
    Intended target audience of the event type, analogue to audience definition for REST APIs
    in rule #219 -- see https://opensource.zalando.com/restful-api-guidelines/#219
owning_application:
  description: |
    Name of the application (eg, as would be used in infrastructure
    application or service registry) owning this `EventType`.
  type: string
  example: price-service
category:
  description: Defines the category of this EventType.
  type: string
  x-extensible-enum:
    - data
    - general
compatibility_mode:
  description: |
    The compatibility mode to evolve the schema.
  type: string
  x-extensible-enum:
    - compatible
    - forward
    - none
  default: forward
schema:
  description: The most recent payload schema for this EventType.
  type: object
  properties:
    version:
      description: Values are based on semantic versioning (eg "1.2.1").
      type: string
      default: '1.0.0'
    created_at:
      description: Creation timestamp of the schema.
      type: string
      readOnly: true
      format: date-time
      example: '1996-12-19T16:39:57-08:00'
```

```
      type:
        description: |
           The schema language of schema definition. Currently only
           json_schema (JSON Schema v04) syntax is defined, but in the
           future there could be others.
        type: string
        x-extensible-enum:
          - json_schema
      schema:
        description: |
            The schema as string in the syntax defined in the field type.
        type: string
    required:
      - type
      - schema
  ordering_key_fields:
    type: array
    description: |
      Indicates which field is used for application level ordering of events.
      It is typically a single field, but also multiple fields for compound
      ordering key are supported (first item is most significant).

      This is an informational only event type attribute for specification of
      application level ordering. Nakadi transportation layer is not affected,
      where events are delivered to consumers in the order they were published.

      Scope of the ordering is all events (of all partitions), unless it is
      restricted to data instance scope in combination with
      `ordering_instance_ids` attribute below.

      This field can be modified at any moment, but event type owners are
      expected to notify consumer in advance about the change.

      *Background:* Event ordering is often created on application level using
      ascending counters, and data providers/consumers do not need to rely on the
      event publication order. A typical example are data instance change events
      used to keep a data store replica in sync. Here you have an order
      defined per instance using data object change counters (aka row update
      version) and the order of event publication is not relevant, because
      consumers for data synchronization skip older instance versions when they
      reconstruct the data object replica state.

    items:
      type: string
      description: |
        Indicates a single ordering field. This is a JsonPointer, which is applied
        onto the whole event object, including the contained metadata and data (in
        case of a data change event) objects. It must point to a field of type
        string or number/integer (as for those the ordering is obvious).

        Indicates a single ordering field. It is a simple path (dot separated) to
        the JSON leaf element of the whole event object, including the contained metadata and data (in
```

```
        case of a data change event) objects. It must point to a field of type
        string or number/integer (as for those the ordering is obvious), and must be
        present in the schema.
      example: "data.order_change_counter"
  ordering_instance_ids:
    type: array
    description: |
      Indicates which field represents the data instance identifier and scope in
      which ordering_key_fields provides a strict order. It is typically a single
      field, but multiple fields for compound identifier keys are also supported.

      This is an informational only event type attribute without specific Nakadi
      semantics for specification of application level ordering. It only can be
      used in combination with `ordering_key_fields`.

      This field can be modified at any moment, but event type owners are expected
      to notify consumer in advance about the change.
    items:
      type: string
      description: |
        Indicates a single key field. It is a simple path (dot separated) to the JSON
        leaf element of the whole event object, including the contained metadata and
        data (in case of a data change event) objects, and it must be present in the
        schema.
     example: "data.order_number"
  created_at:
    description: When this event type was created.
    type: string
    pattern: date-time
  updated_at:
    description: When this event type was last updated.
    type: string
    pattern: date-time
```

APIs such as registries supporting event types, may extend the model, including the set of supported categories and schema formats. For example the Nakadi API's event category registration also allows the declaration of validation and enrichment strategies for events, along with supplemental information, such as how events are partitioned in the stream (see **SHOULD** use the hash partition strategy for data change events).

MUST follow naming convention for event type names [213]

Event type names must (or should, see **MUST**/**SHOULD**/**MAY** use functional naming schema for details and definition) conform to the functional naming depending on the audience as follows:

```
<event-type-name>       ::= <functional-event-name> | <application-event-name>

<functional-event-name> ::= <functional-name>.<event-name>[.<version>]
```

```
<event-name>            ::= [a-z][a-z0-9-]* -- free event name (functional name)

<version>               ::= [Vv][0-9.]* -- major version of non compatible schemas
```

Hint: The following convention (e.g. used by legacy STUPS infrastructure) is deprecated and **only** allowed for internal event type names:

```
<application-event-name> ::= [<organization-id>.]<application-id>.<event-name>
<organization-id>  ::= [a-z][a-z0-9-]* -- organization identifier, e.g. team identifier
<application-id>   ::= [a-z][a-z0-9-]* -- application identifier
```

Note: consistent naming should be used whenever the same entity is exposed by a data change event and a RESTful API.

MUST indicate ownership of event types [207]

Event definitions must have clear ownership - this can be indicated via the `owning_application` field of the EventType.

Typically there is one producer application, which owns the EventType and is responsible for its definition, akin to how RESTful API definitions are managed. However, the owner may also be a particular service from a set of multiple services that are producing the same kind of event.

MUST carefully define the compatibility mode [245]

Event type owners must pay attention to the choice of compatibility mode. The mode provides a means to evolve the schema. The range of modes are designed to be flexible enough so that producers can evolve schemas while not inadvertently breaking existing consumers:

- `none` : Any schema modification is accepted, even if it might break existing producers or consumers. When validating events, undefined properties are accepted unless declared in the schema.
- `forward` : A schema `S1` is forward compatible if the previously registered schema, `S0` can read events defined by `S1` - that is, consumers can read events tagged with the latest schema version using the previous version as long as consumers follow the robustness principle described in the guideline's API design principles.
- `compatible` : This means changes are fully compatible. A new schema, `S1` , is fully compatible when every event published since the first schema version will validate against the latest schema. In compatible mode, only the addition of new optional properties and

definitions to an existing schema is allowed. Other changes are forbidden.

MUST ensure event schema conforms to OpenAPI schema object [196]

To align the event schema specifications to API specifications, we use the Schema Object as defined by the OpenAPI Specifications to define event schemas. This is particularly useful for events that represent data changes about resources also used in other APIs.

The OpenAPI Schema Object is an **extended subset** of JSON Schema Draft 4. For convenience, we highlight some important differences below. Please refer to the OpenAPI Schema Object specification for details.

As the OpenAPI Schema Object specification *removes* some JSON Schema keywords, the following properties **must not** be used in event schemas:

- `additionalItems`
- `contains`
- `patternProperties`
- `dependencies`
- `propertyNames`
- `const`
- `not`
- `oneOf`

On the other side Schema Object *redefines* some JSON Schema keywords:

- `additionalProperties` : For event types that declare compatibility guarantees, there are recommended constraints around the use of this field. See the guideline **SHOULD** avoid `additionalProperties` in event type schemas for details.

Finally, the Schema Object *extends* JSON Schema with some keywords:

- `readOnly` : events are logically immutable, so `readOnly` can be considered redundant, but harmless.
- `discriminator` : to support polymorphism, as an alternative to `oneOf` .

- `^x-` : patterned objects in the form of vendor extensions can be used in event type schema, but it might be the case that general purpose validators do not understand them to enforce a validation check, and fall back to must-ignore processing. A future version of the guidelines may define well known vendor extensions for events.

SHOULD avoid `additionalProperties` in event type schemas [210]

Event type schema should avoid using `additionalProperties` declarations, in order to support schema evolution.

Events are often intermediated by publish/subscribe systems and are commonly captured in logs or long term storage to be read later. In particular, the schemas used by publishers and consumers can
drift over time. As a result, compatibility and extensibility issues that happen less frequently with client-server style APIs become important and regular considerations for event design. The guidelines recommend the following to enable event schema evolution:

- Publishers who intend to provide compatibility and allow their schemas to evolve safely over time **must not** declare an `additionalProperties` field with a value of `true` (i.e., a wildcard extension point). Instead they must define new optional fields and update their schemas in advance of publishing those fields.
- Consumers **must** ignore fields they cannot process and not raise errors. This can happen if they are processing events with an older copy of the event schema than the one containing the new definitions specified by the publishers.

The above constraint does not mean fields can never be added in future revisions of an event type schema - additive compatible changes are allowed, only that the new schema for an event type must define the field first before it is published within an event. By the same turn the consumer must ignore fields it does not know about from its copy of the schema, just as they would as an API client - that is, they cannot treat the absence of an `additionalProperties` field as though the event type schema was closed for extension.

Requiring event publishers to define their fields ahead of publishing avoids the problem of *field redefinition*. This is when a publisher defines a field to be of a different type that was already being emitted, or, is changing the type of an undefined field. Both of these are prevented by not using `additionalProperties` .

See also rule **MUST** treat OpenAPI specification as open for extension by default in the REST Design - Compatibility section for further guidelines on the use of `additionalProperties`.

MUST use semantic versioning of event type schemas [246]

Event schemas must be versioned — analog to **MUST** use semantic versioning for REST API definitions. The compatibility mode interact with revision numbers in the schema `version` field, which follows semantic versioning (MAJOR.MINOR.PATCH):

- Changing an event type with compatibility mode `compatible` or `forward` can lead to a PATCH or MINOR version revision. MAJOR breaking changes are not allowed.
- Changing an event type with compatibility mode `none` can lead to PATCH, MINOR or MAJOR level changes.

The following examples illustrate these relations:

- Changes to the event type's `title` or `description` are considered PATCH level.
- Adding new optional fields to an event type's schema is considered a MINOR level change.
- All other changes are considered MAJOR level, such as renaming or removing fields, or adding new required fields.

---

19. EVENT Basics - Event Categories

An *event category* describes a generic class of event types. The guidelines define two such categories:

- General Event: a general purpose category.
- Data Change Event: a category to inform about data entity changes and used e.g. for data replication based data integration.

MUST ensure that events conform to an event category [198]

A category describes a predefined structure (e.g. including event metadata as part of the event payload) that event publishers must conform to along with standard information specific for the event category (e.g. the operation for data change events).

The general event category

The structure of the *General Event Category* is shown below as an OpenAPI Schema Object definition:

```
GeneralEvent:
  description: |
    A general kind of event. Event kinds based on this event define their
    custom schema payload as the top level of the document, with the
    "metadata" field being required and reserved for standard metadata. An
    instance of an event based on the event type thus conforms to both the
    EventMetadata definition and the custom schema definition.
    Hint: In earlier versions this category was called the Business Category.
  required:
    - metadata
  properties:
    metadata:
        $ref: '#/definitions/EventMetadata'
```

Event types based on the General Event Category define their custom schema payload at the top-level of the document, with the metadata field being reserved for standard information (the contents of metadata are described further down in this section).

Note:

- The General Event was called a *Business Event* in earlier versions of the guidelines. Implementation experience has shown that the category's structure gets used for other kinds of events, hence the name has been generalized to reflect how teams are using it.
- The General Event is still useful and recommended for the purpose of defining events that drive a business process.
- The Nakadi broker still refers to the General Category as the Business Category and uses the keyword `business` for event type registration. Other than that, the JSON structures are identical.

See **MUST** use general events to signal steps in business processes for more guidance on how to use the category.

The data change event category

The *Data Change Event Category* structure is shown below as an OpenAPI Schema Object:

```
DataChangeEvent:
  description: |
```

```
    Represents a change to an entity. The required fields are those
    expected to be sent by the producer, other fields may be added
    by intermediaries such as a publish/subscribe broker. An instance
    of an event based on the event type conforms to both the
    DataChangeEvent's definition and the custom schema definition.
  required:
    - metadata
    - data_op
    - data_type
    - data
  properties:
    metadata:
      description: The metadata for this event.
      $ref: '#/definitions/EventMetadata'
    data:
      description: |
        Contains custom payload for the event type. The payload must conform
        to a schema associated with the event type declared in the metadata
        object's `event_type` field.
      type: object
    data_type:
      description: name of the (business) data entity that has been mutated
      type: string
      example: 'sales_order.order'
    data_op:
      type: string
      enum: ['C', 'U', 'D', 'S']
      description: |
        The type of operation executed on the entity:

        - C: Creation of an entity
        - U: An update to an entity.
        - D: Deletion of an entity.
        - S: A snapshot of an entity at a point in time.
```

The Data Change Event Category is structurally different to the General Event Category by defining a field called **`data`** as container for the custom payload, as well as specific information related to data changes in the **`data_op`** .

The following guidelines specifically apply to Data Change Events:

- **MUST** use data change events to signal mutations
- **MUST** provide explicit event ordering for data change events
- **SHOULD** ensure that data change events match the APIs resources
- **SHOULD** use the hash partition strategy for data change events

Event metadata

MUST provide mandatory event metadata [247]

The General and Data Change event categories share a common structure for *metadata* representing generic event information. Parts of the metadata is provided by the Nakadi event messaging platform, but event identifier (eid) and event creation timestamp (occurred_at) have to be provided by the event producers. The metadata structure is defined below as an OpenAPI Schema Object:

```
EventMetadata:
  type: object
  description: |
    Carries metadata for an Event along with common fields. The required
    fields are those expected to be sent by the producer, other fields may be
    added by intermediaries such as publish/subscribe broker.
  required:
    - eid
    - occurred_at
  properties:
    eid:
      description: Identifier of this event.
      type: string
      format: uuid
      example: '105a76d8-db49-4144-ace7-e683e8f4ba46'
    event_type:
      description: The name of the EventType of this Event.
      type: string
      example: 'example.important-business-event'
    occurred_at:
      description: |
         Technical timestamp of when the event object was created during processing
         of the business event by the producer application. Note, it may differ from
         the timestamp when the related real-world business event happened (e.g. when
         the packet was handed over to the customer), which should be passed separately
         via an event type specific attribute.
         Depending on the producer implementation, the timestamp is typically some
         milliseconds earlier than when the event is published and received by the
         API event post endpoint server -- see below.
      type: string
      format: date-time
      example: '1996-12-19T16:39:57-08:00'
    received_at:
      description: |
        Timestamp of when the event was received via the API event post endpoints.
        It is automatically enriched, and events will be rejected if set by the
        event producer.
      type: string
      readOnly: true
```

```
      format: date-time
      example: '1996-12-19T16:39:57-08:00'
    version:
      description: |
        Version of the schema used for validating this event. This may be
        enriched upon reception by intermediaries. This string uses semantic
        versioning.
      type: string
      readOnly: true
    parent_eids:
      description: |
        Event identifiers of the Event that caused the generation of
        this Event. Set by the producer.
      type: array
      items:
        type: string
        format: uuid
      example: '105a76d8-db49-4144-ace7-e683e8f4ba46'
    flow_id:
      description: |
        A flow-id for this event (corresponds to the X-Flow-Id HTTP header).
      type: string
      example: 'JAh6xH4OQhCJ9PutIV_RYw'
    partition:
      description: |
        Indicates the partition assigned to this Event. Used for systems
        where an event type's events can be sub-divided into partitions.
      type: string
      example: '0'
```

Please note that intermediaries acting between the producer of an event and its ultimate consumers, may perform operations like validation of events and enrichment of an event's metadata. For example brokers such as Nakadi, can validate and enrich events with arbitrary additional fields that are not specified here and may set default or other values, if some of the specified fields are not supplied. How such systems work is outside the scope of these guidelines but producers and consumers working with such systems should look into their documentation for additional information.

MUST provide unique event identifiers [211]

Event publishers must provide the `eid` (event identifier) as a standard event object field and part of the event metadata. The `eid` must be a unique identifier for the event in the scope of the event type / stream lifetime.

Event producers must use the same `eid` when publishing the same event object multiple times. For instance, producers must ensure that publish event retries — e.g. due to temporary Nakadi or

network failures or fail-overs — use the same `eid` value as the initial (possibly failed) attempt.

The `eid` supports event consumers in detecting and being robust against event duplicates — see **MUST** be robust against duplicates when consuming events.

Hint: Using the same eid for retries can be ensured, for instance, by generating a UUID either (i) as part of the event object construction and using some form of intermediate persistence, like an event publish retry queue, or (ii) via a deterministic function that computes the UUID value from the event fields as input (without random salt).

MUST use general events to signal steps in business processes [201]

When publishing events that represent steps in a business process, event types **must** be based on the General Event category. All your events of a single business process will conform to the following rules:

- Business events must contain a specific identifier field (a business process id or "bp-id") similar to flow-id to allow for efficient aggregation of all events in a business process execution.
- Business events must contain a means to correctly order events in a business process execution. In distributed settings where monotonically increasing values (such as a high precision timestamp that is assured to move forwards) cannot be obtained, the `parent_eids` data structure allows causal relationships to be declared between events.
- Business events should only contain information that is new to the business process execution at the specific step/arrival point.
- Each business process sequence should be started by a business event containing all relevant context information.
- Business events must be published reliably by the service.

At the moment we cannot state whether it's best practice to publish all the events for a business process using a single event type and represent the specific steps with a state field, or whether to use multiple event types to represent each step. For now we suggest assessing each option and sticking to one for a given business process.

SHOULD provide explicit event ordering for general events [203]

Event processing consumer applications need the order information to reconstruct the business event stream, for instance, in order to replay events in error situations, or to execute analytical use cases outside the context of the original event stream consumption. All general events (fka business events) **should** be provided with the explicit information about the business ordering of the events. To accomplish this event ordering the event type definition:

- **must** specify the `ordering_key_fields` property to indicate which field(s) contain the ordering key, and
- **should** specify the `ordering_instance_ids` property to define which field(s) represents the business entity instance identifier.

Note: The `ordering_instance_ids` restrict the scope in which the `ordering_key_fields` provide the strict order. If undefined, the ordering is assumed to be provided in scope of all events.

The business ordering information can be provided – among other ways – by maintaining…

- a strictly monotonically increasing version of entity instances (e.g. created as row update counter by a database), or
- a strictly monotonically increasing sequence counter (maintained per partition or event type).

Hint: timestamps are often a bad choice, since in distributed systems events may occur at the same time, or clocks are not exactly synchronized, or jump forward and backward to compensate drifts or leap-seconds. If you use anyway timestamps to indicate event ordering, you *must* carefully ensure that the designated event order is not messed up by these effects and use UTC time zone format.

Note: The `received_at` and `partition_offset` metadata set by Nakadi typically is different from the business event ordering, since (1) Nakadi is a distributed concurrent system without atomic, ordered event creation and (2) the application's implementation of event publishing may not exactly reflect the business order. The business ordering information is application knowledge, and implemented in the scope of event partitions or specific entities, but may also comprise all events, if scaling requirements are low.

MUST use data change events to signal mutations [202]

You **must** use data change events to signal changes of stored entity instances and facilitate e.g. change data capture (CDC). Event sourced change data capture is crucial for our data integration

architecture as it supports the logical replication (and reconstruction) of the application datastores to the data analytics and AI platform as transactional source datasets.

- Change events must be provided when publishing events that represent created, updated, or deleted data.
- Change events must provide the complete entity data including the identifier of the changed instance to allow aggregation of all related events for the entity.
- Change events **MUST** provide explicit event ordering for data change events.
- Change events must be published reliably by the service.

MUST provide explicit event ordering for data change events [242]

While the order information is recommended for business events, it **must** be provided for data change events. The ordering information defines the (create, update, delete) change order of the data entity instances managed via the application's transactional datastore. It is needed for change data capture to keep transactional dataset replicas in sync as source for data analytics.

For details about how to provide the data change ordering information, please check **SHOULD** provide explicit event ordering for general events.

Exception: In situations where the transactional data is 'append only', i.e. entity instances are only created, but never updated or deleted, the ordering information may not be provided.

SHOULD use the hash partition strategy for data change events [204]

The `hash` partition strategy allows a producer to define which fields in an event are used as input to compute a logical partition the event should be added to. Partitions are useful as they allow supporting systems to scale their throughput while provide local ordering for event entities.

The `hash` option is particularly useful for data changes as it allows all related events for an entity to be consistently assigned to a partition, providing a relative ordered stream of events for that entity. This is because while each partition has a total ordering, ordering across partitions is not assured by a supporting system, thus it is possible for events sent across partitions to appear in a different order to consumers that the order they arrived at the server.

When using the `hash` strategy the partition key in almost all cases should represent the entity

being changed and not a per event or change identifier such as the `eid` field or a timestamp. This ensures data changes arrive at the same partition for a given entity and can be consumed effectively by clients.

There may be exceptional cases where data change events could have their partition strategy set to be the producer defined or random options, but generally `hash` is the right option - that is while the guidelines here are a "should", they can be read as "must, unless you have a very good reason".

---

20. EVENT Design

SHOULD avoid writing sensitive data to events [200]

Event data security is supported by Nakadi Event Bus mechanisms for access control and authorization of publishing or consuming events. However, we avoid writing sensitive data (e.g. personal data like e-mail or address) to events unless it is needed for the business. Sensitive data create additional obligations for access control and compliance and generally increases data protection risks.

MUST be robust against duplicates when consuming events [214]

Duplicate events are multiple occurrences of the same event object representing the same (business or data change) event instance.

Event consumers must be robust against duplicate events. Depending on the use case, being robust implies that event consumers need to *deduplicate events*, i.e. to ignore duplicates in event processing. For instance, for accounting reporting a high accuracy is required by the business, and duplicates need to be explicitly ignored, whereas customer behavior reporting (click rates) might not care about event duplicates since accuracy in per-mille range is not needed.

Hint: Event consumers are supported in deduplication:

- Deduplication can be based on the `eid` as mandatory standard for all events — see **MUST** provide unique event identifiers.
- Processing data change events to replay data changes and keep transactional data copies in sync (CDC) is robust against duplicates because it is based on data keys and change ordering — see **MUST** use data change events to signal mutations and **MUST** provide explicit

event ordering for data change events.

- Data analytics users of the Data Lake are well advised to use curated data as a source for analytics. Raw event datasets materialized in the lake are typically cleaned-up (including deduplication and data synchronization) and provided as transactional data copy or curated data products, for instance curated product data [internal link] or curated sales data [internal link].

Context: One *source of duplicate events* are the event producers, for instance, due to publish call retries or publisher client failovers. Another source is Nakadi's 'at-least-once' delivery promise (like provided by most message broker distributed systems). It also currently applies to the Data Lake event materialization as raw event datasets (in Delta or JSON format) for Data Analytics. From an event consumer point of view, duplicate events may also be created when consumers reprocess parts of the event stream, for instance, due to inaccurate continuation after failures. Event publishers and infrastructure systems should keep event duplication at a minimum typically below per-mille range. (In Nov. 2022, for instance, we observed <0.2 ‰ daily event duplicate rate (95th percentile) for high volume events.)

SHOULD design for idempotent out-of-order processing [212]

Events that are designed for idempotent out-of-order processing allow for extremely resilient systems: If processing an event fails, consumers and producers can skip/delay/retry it without stopping the world or corrupting the processing result.

To enable this freedom of processing, you must explicitly design for idempotent out-of-order processing: Either your events must contain enough information to infer their original order during consumption or your domain must be designed in a way that order becomes irrelevant.

As common example similar to data change events, idempotent out-of-order processing can be supported by sending the following information:

- the process/resource/entity identifier,
- a monotonically increasing ordering key and
- the process/resource state after the change.

A receiver that is interested in the current state can then ignore events that are older than the last processed event of each resource. A receiver interested in the history of a resource can use the ordering key to recreate a (partially) ordered sequence of events.

MUST ensure that events define useful business resources [199]

Events are intended to be used by other services including business process/data analytics and monitoring. They should be based around the resources and business processes you have defined for your service domain and adhere to its natural lifecycle (see also **SHOULD** model complete business processes and **SHOULD** define *useful* resources).

As there is a cost in creating an explosion of event types and topics, prefer to define event types that are abstract/generic enough to be valuable for multiple use cases, and avoid publishing event types without a clear need.

SHOULD ensure that data change events match the APIs resources [205]

A data change event's representation of an entity should correspond to the REST API representation.

There's value in having the fewest number of published structures for a service. Consumers of the service will be working with fewer representations, and the service owners will have less API surface to maintain. In particular, you should only publish events that are interesting in the domain and abstract away from implementation or local details - there's no need to reflect every change that happens within your system.

There are cases where it could make sense to define data change events that don't directly correspond to your API resource representations. Some examples are -

- Where the API resource representations are very different from the datastore representation, but the physical data are easier to reliably process for data integration.
- Publishing aggregated data. For example a data change to an individual entity might cause an event to be published that contains a coarser representation than that defined for an API
- Events that are the result of a computation, such as a matching algorithm, or the generation of enriched data, and which might not be stored as entity by the service.

MUST maintain backwards compatibility for events [209]

Changes to events must be based around making additive and backward compatible changes. This follows the **MUST** not break backward compatibility guideline.

In the context of events, compatibility issues are complicated by the fact that producers and consumers of events are highly asynchronous and can't use content-negotiation techniques that are available to REST style clients and servers. This places a higher bar on producers to maintain compatibility as they will not be in a position to serve versioned media types on demand.

For event schema, these are considered backward compatible changes, as seen by consumers -

- Adding new optional fields to JSON objects.
- Changing the order of fields (field order in objects is arbitrary).
- Changing the order of values with same type in an array.
- Removing optional fields.
- Removing an individual value from an enumeration.
- Adding new value to a `x-extensible-enum` field (see rule 112 and rule 108).

These are considered backwards-incompatible changes, as seen by consumers -

- Removing required fields from JSON objects.
- Changing the default value of a field.
- Changing the type of a field, object, enum or array.
- Changing the order of values with different type in an array (also known as a tuple).
- Adding a new optional field to redefine the meaning of an existing field (also known as a co-occurrence constraint).
- Adding a value to an enumeration. Instead, you **SHOULD** use open-ended list of values (`x-extensible-enum`) for enumeration types.

Appendix A: References

This section collects links to documents to which we refer, and base our guidelines on.

OpenAPI specification

- **OpenAPI specification**
- **OpenAPI specification mind map**

Publications, specifications and standards

- **RFC 3339**: Date and Time on the Internet: Timestamps
- **RFC 4122**: A Universally Unique IDentifier (UUID) URN Namespace
- **RFC 4627**: The application/json Media Type for JavaScript Object Notation (JSON)
- **RFC 4648**: The Base16, Base32, and Base64 Data Encodings
- **RFC 6585**: Additional HTTP Status Codes
- **RFC 6902**: JavaScript Object Notation (JSON) Patch
- **RFC 7159**: The JavaScript Object Notation (JSON) Data Interchange Format
- **RFC 7240**: Prefer Header for HTTP
- **RFC 7396**: JSON Merge Patch
- **RFC 8288**: Web Linking
- **RFC 9110**: HTTP Semantics
- **RFC 9111**: HTTP Caching
- **RFC 9457**: Problem Details for HTTP APIs
- **ISO 8601**: Date and time format
- **ISO 3166-1 alpha-2**: Two letter country codes
- **ISO 639-1**: Two letter language codes
- **ISO 4217**: Currency codes
- **BCP 47**: Tags for Identifying Languages

Dissertations

- **Roy Thomas Fielding - Architectural Styles and the Design of Network-Based Software Architectures**: This is the text which defines what REST is.

Books

- **REST in Practice: Hypermedia and Systems Architecture**
- **Build APIs You Won't Hate**
- **InfoQ eBook - Web APIs: From Start to Finish**

Blogs

- **Lessons-learned blog: Thoughts on RESTful API Design**

Appendix B: Tooling

This is not a part of the actual guidelines, but might be helpful for following them. Using a tool mentioned here doesn't automatically ensure you follow the guidelines.

API first integrations

The following frameworks were specifically designed to support the API First workflow with OpenAPI YAML files (sorted alphabetically):

- **Connexion**: OpenAPI First framework for Python on top of Flask
- **Api-First-Hand**: API-First Play Bootstrapping Tool for Swagger/OpenAPI specs
- **Swagger Codegen**: template-driven engine to generate client code in different languages by parsing Swagger Resource Declaration
- **Swagger Codegen Tooling**: plugin for Maven that generates pieces of code from OpenAPI specification
- **Swagger Plugin for IntelliJ IDEA**: plugin to help you easily edit Swagger specification files inside IntelliJ IDEA

The Swagger/OpenAPI homepage lists more Community-Driven Language Integrations, but most of them do not fit our API First approach.

Support libraries

These utility libraries support you in implementing various parts of our RESTful API guidelines (sorted alphabetically):

- **Problem**: Java library that implements application/problem+json
- **Problems for Spring Web MVC**: library for handling Problems in Spring Web MVC
- **Jackson Datatype Money**: extension module to properly support datatypes of javax.money
- **Tracer**: call tracing and log correlation in distributed systems

Appendix C: Best practices

The best practices presented in this section are not part of the actual guidelines, but should provide guidance for common challenges we face when implementing RESTful APIs.

Cursor-based pagination in RESTful APIs

Cursor-based pagination is a very powerful and valuable technique (see also **SHOULD** prefer cursor-based pagination, avoid offset-based pagination) that allows to efficiently provide a stable view on changing data. This is obtained by using an anchor element that allows to retrieve all page elements directly via an ordering combined-index, usually based on `created_at` or `modified_at`. Simple said, the cursor is the information set needed to reconstruct the database query that retrieves the minimal page information from the data storage.

The `cursor` itself is an opaque string, transmitted forth and back between service and clients, that must never be **inspected** or **constructed** by clients. Therefore, it is good practice to encode (encrypt) its content in a non-human-readable form.

The `cursor` content usually consists of a pointer to the anchor element defining the page position in the collection, a flag whether the element is included or excluded into/from the page, the retrieval direction, and a hash over the applied query filters (or the query filter itself) to safely re-create the collection. It is important to note, that a `cursor` should be always defined in relation to the current page to anticipate all occurring changes when progressing.

The `cursor` is usually defined as an encoding of the following information:

```
Cursor:
  descriptions: >
    Cursor structure that contains all necessary information to efficiently
    retrieve a page from the data store.
  type: object
  properties:
    position:
      description: >
        Object containing the keys pointing to the anchor element that is
        defining the collection resource page. Normally the position is given
        by the first or the last page element. The position object contains all
        values required to access the element efficiently via the ordered,
        combined index, e.g `modified_at`, `id`.
      type: object
      properties: ...

    element:
      description: >
        Flag whether the anchor element, which is pointed to by the `position`,
        should be *included* or *excluded* from the result set. Normally, only
```

```
      the current page includes the pointed to element, while all others are
      exclude it.
    type: string
    enum: [ INCLUDED, EXCLUDED ]

  direction:
    description: >
      Flag for the retrieval direction that is defining which elements to
      choose from the collection resource starting from the anchor elements
      position. It is either *ascending* or *descending* based on the
      ordering combined index.
    type: string
    enum: [ ASCENDING, DESCENDING ]

  query_hash:
    description: >
      Stable hash calculated over all query filters applied to create the
      collection resource that is represented by this cursor.
    type: string

  query:
    description: >
      Object containing all query filters applied to create the collection
      resource that is represented by this cursor.
    type: object
    properties: ...

required:
  - position
  - element
  - direction
```

Note: In case of complex and long search requests, e.g. when `GET with body` is already required, the `cursor` may not be able to include the `query` because of common HTTP parameter size restrictions. In this case the `query` filters should be transported via body - in the request as well as in the response, while the pagination consistency should be ensured via the `query_hash` .

Remark: It is also important to check the efficiency of the data-access. You need to make sure that you have a fully ordered stable index, that allows to efficiently resolve all elements of a page. If necessary, you need to provide a combined index that includes the `id` to ensure the full order and additional filter criteria to ensure efficiency.

Further reading

- Twitter
- Use the Index, Luke

- Paging in PostgreSQL

Optimistic locking in RESTful APIs

Introduction

Optimistic locking might be used to avoid concurrent writes on the same entity, which might cause data loss. A client always has to retrieve a copy of an entity first and specifically update this one. If another version has been created in the meantime, the update should fail. In order to make this work, the client has to provide some kind of version reference, which is checked by the service, before the update is executed. Please read the more detailed description on how to update resources via `PUT` in the HTTP Requests Section.

A RESTful API usually includes some kind of search endpoint, which will then return a list of result entities. There are several ways to implement optimistic locking in combination with search endpoints which, depending on the approach chosen, might lead to performing additional requests to get the current version of the entity that should be updated.

`ETag` with `If-Match` header

An `ETag` can only be obtained by performing a `GET` request on the single entity resource before the update, i.e. when using a search endpoint an additional request is necessary.

Example:

```
< GET /orders

> HTTP/1.1 200 OK
> {
>   "items": [
>     { "id": "O0000042" },
>     { "id": "O0000043" }
>   ]
> }

< GET /orders/BO0000042

> HTTP/1.1 200 OK
> ETag: osjnfkjbnkq3jlnksjnvkjlsbf
> { "id": "BO0000042", ... }

< PUT /orders/O0000042
< If-Match: osjnfkjbnkq3jlnksjnvkjlsbf
< { "id": "O0000042", ... }
```

```
> HTTP/1.1 204 No Content
```

Or, if there was an update since the `GET` and the entity's `ETag` has changed:

```
> HTTP/1.1 412 Precondition failed
```

Pros

- RESTful solution

Cons

- Many additional requests are necessary to build a meaningful front-end

`ETags` in result entities

The ETag for every entity is returned as an additional property of that entity. In a response containing multiple entities, every entity will then have a distinct `ETag` that can be used in subsequent `PUT` requests.

In this solution, the `etag` property should be `readonly` and never be expected in the `PUT` request payload.

Example:

```
< GET /orders

> HTTP/1.1 200 OK
> {
>   "items": [
>     { "id": "O0000042", "etag": "osjnfkjbnkq3jlnksjnvkjlsbf", "foo": 42, "bar": true },
>     { "id": "O0000043", "etag": "kjshdfknjqlowjdsljdnfkjbkn", "foo": 24, "bar": false }
>   ]
> }

< PUT /orders/O0000042
< If-Match: osjnfkjbnkq3jlnksjnvkjlsbf
< { "id": "O0000042", "foo": 43, "bar": true }

> HTTP/1.1 204 No Content
```

Or, if there was an update since the `GET` and the entity's `ETag` has changed:

```
> HTTP/1.1 412 Precondition failed
```

Pros

- Perfect optimistic locking

Cons

- Information that only belongs in the HTTP header is part of the business objects

Version numbers

The entities contain a property with a version number. When an update is performed, this version number is given back to the service as part of the payload. The service performs a check on that version number to make sure it was not incremented since the consumer got the resource and performs the update, incrementing the version number.

Since this operation implies a modification of the resource by the service, a **`POST`** operation on the exact resource (e.g. **`POST /orders/O0000042`**) should be used instead of a **`PUT`** .

In this solution, the **`version`** property is not **`readonly`** since it is provided at **`POST`** time as part of the payload.

Example:

```
< GET /orders

> HTTP/1.1 200 OK
> {
>   "items": [
>     { "id": "O0000042", "version": 1,  "foo": 42, "bar": true },
>     { "id": "O0000043", "version": 42, "foo": 24, "bar": false }
>   ]
> }

< POST /orders/O0000042
< { "id": "O0000042", "version": 1, "foo": 43, "bar": true }

> HTTP/1.1 204 No Content
```

or if there was an update since the **`GET`** and the version number in the database is higher than the one given in the request body:

```
> HTTP/1.1 409 Conflict
```

Pros

- Perfect optimistic locking

Cons

- Functionality that belongs into the HTTP header becomes part of the business object
- Using `POST` instead of PUT for an update logic (not a problem in itself, but may feel unusual for the consumer)

`Last-Modified` / `If-Unmodified-Since`

In HTTP 1.0 there was no `ETag` and the mechanism used for optimistic locking was based on a date. This is still part of the HTTP protocol and can be used. Every response contains a `Last-Modified` header with a HTTP date. When requesting an update using a `PUT` request, the client has to provide this value via the header `If-Unmodified-Since`. The server rejects the request, if the last modified date of the entity is after the given date in the header.

This effectively catches any situations where a change that happened between `GET` and `PUT` would be overwritten. In the case of multiple result entities, the `Last-Modified` header will be set to the latest date of all the entities. This ensures that any change to any of the entities that happens between `GET` and `PUT` will be detectable, without locking the rest of the batch as well.

Example:

```
< GET /orders

> HTTP/1.1 200 OK
> Last-Modified: Wed, 22 Jul 2009 19:15:56 GMT
> {
>   "items": [
>     { "id": "O0000042", ... },
>     { "id": "O0000043", ... }
>   ]
> }

< PUT /block/O0000042
< If-Unmodified-Since: Wed, 22 Jul 2009 19:15:56 GMT
< { "id": "O0000042", ... }

> HTTP/1.1 204 No Content
```

Or, if there was an update since the `GET` and the entities last modified is later than the given date:

```
> HTTP/1.1 412 Precondition failed
```

Pros

- Well established approach that has been working for a long time
- No interference with the business objects; the locking is done via HTTP headers only
- Very easy to implement
- No additional request needed when updating an entity of a search endpoint result

Cons

- If a client communicates with two different instances and their clocks are not perfectly in sync, the locking could potentially fail

Conclusion

We suggest to either use the `ETag` *in result entities* or `Last-Modified` / `If-Unmodified-Since` approach.

Handling compatible API extensions

In the clients

Client must not eliminate unknown optional fields from the fetched resource payload, and to serialize them later when submitting the complete resource payload back to the API server via `PUT` (**MUST** prepare clients to accept compatible API extensions).

When using Java with Jackson serialization, for example, that can be achieved by including a field in the Java class representing the API resource, like the following one:

```
@JsonAnyGetter
@JsonAnySetter
private Map<String, JsonNode> additionalProperties = new HashMap<>();
```

Appendix D: Changelog

This change log only contains major changes and lists major changes since October 2016.

Non-major changes are editorial-only changes or minor changes of existing guidelines, e.g.

adding new error code or specific example. Major changes are changes that come with additional obligations, or even change an existing guideline obligation. Major changes are listed as "Rule Changes" below.

Hint: Most recent major changes might be missing in the list since we update it only occasionally, not with each pull request, to avoid merge commits. Please have a look at the commit list in Github to see a list of all changes.

Rule Changes

- `2024-06-27` : Clarified usage of `x-extensible-enum` for events in **SHOULD** use open-ended list of values (`x-extensible-enum`) for enumeration types. #807
- `2024-06-25` : Relaxed naming convention for date/time properties in **SHOULD** use naming convention for date/time properties. #811
- `2024-06-11` : Linked **SHOULD** use standard formats for time duration and interval properties (duration / period) from **MUST** use standard data formats. #810
- `2024-05-06` : Added new rule **SHOULD** select appropriate one of date or date-time format on selecting appropriate date or date-time format. #808
- `2024-04-16` : Removed sort example from simple query language in **SHOULD** design simple query languages using query parameters. Enhanced clarity for 'uid' usage and permission naming convention in **MUST** define and assign permissions (scopes) and **MUST** follow the naming convention for permissions (scopes). #804 #801
- `2024-03-21` : Added best practices section for **MUST** prepare clients to accept compatible API extensions on handling compatible API extensions. #799
- `2024-03-05` : Updated security section about uid scope in **MUST** define and assign permissions (scopes). #794
- `2024-02-29` : Improved guidance on `POST` usage in **MUST** use HTTP methods correctly. #791
- `2024-02-21` : Fixed discrepancy between **SHOULD** design APIs conservatively and **MUST** treat OpenAPI specification as open for extension by default regarding handling of unexpected fields. #793
- `2023-12-12` : Improved response code guidance in **SHOULD** only use most common HTTP status codes. #789
- `2023-11-22` : Added new rule **MAY** support asynchronous request processing for supporting asynchronous request processing. #787

- **`2023-07-21`** : Improved guidance on total counts in **SHOULD** avoid a total result count. #731
- **`2023-05-12`** : Added new rule **SHOULD** be aware of services not fully supporting JSON/unicode recommending not to use redirection codes. #762
- **`2023-05-08`** : Improved guidance on sanitizing JSON payload in **SHOULD** be aware of services not fully supporting JSON/unicode. #759
- **`2023-04-18`** : Added new rule **SHOULD** design single resource schema for reading and writing recommending to design single resource schema for reading and writing. Added exception for partner IAM to **MUST** secure endpoints. #764 #767
- **`2022-12-20`** : Clarify that event consumers must be robust against duplicates in **MUST** be robust against duplicates when consuming events. #749
- **`2022-10-18`** : Add X-Zalando-Customer to list of proprietary headers in **SHOULD** use only the specified proprietary Zalando headers. #743
- **`2022-09-21`** : Clarify that functional naming schema in **MUST**/**SHOULD**/**MAY** use functional naming schema is a **MUST**/**SHOULD**/**MAY** rule. #740
- **`2022-07-26`** : Improve guidance for return code usage for (robust) create operations in **MUST** use HTTP methods correctly. #735
- **`2022-07-21`** : Improve format and time interval guidance in **MUST** use standard data formats and **SHOULD** use standard formats for time duration and interval properties. #733
- **`2022-05-24`** : Define `next` page link as optional in [pagination-fields]. #726
- **`2022-04-19`** : Change **SHOULD** avoid writing sensitive data to events from **MUST** to **SHOULD** avoid providing sensitive data with events. #723
- **`2022-03-22`** : More clarity about when error specification definitions can be omitted in **MUST** specify success and error responses. More clarity around HTTP status codes in **MUST** use official HTTP status codes. More clarity for avoiding null for boolean fields in **MUST** not use `null` for boolean properties. #715 #720 #721
- **`2022-01-26`** : Exclude 'type' from common field names in **SHOULD** use naming convention for date/time properties. Improve clarity around usage of extensible enums in **SHOULD** use open-ended list of values (`x-extensible-enum`) for enumeration types. #714 #717
- **`2021-12-22`** : Clarify that event id must not change in retry situations when producers **MUST** provide unique event identifiers. #694
- **`2021-12-09`** : Improve clarity on `PATCH` and `PUT` usage in rule **MUST** use HTTP methods correctly. Only use codes registered via IANA in rule **MUST** use official HTTP status codes.

cb0624b

- **2021-12-09** : event id must not change in retry situations when producers **MUST** provide unique event identifiers.
- **2021-11-24** : restructuring of the document and some rules.
- **2021-10-18** : new rule **SHOULD** use content negotiation, if clients may choose from different resource representations.
- **2021-10-12** : improve clarity on `PATCH` usage in rule **MUST** use HTTP methods correctly.
- **2021-08-24** : improve clarity on `PUT` usage in rule **MUST** use HTTP methods correctly.
- **2021-08-24** : only use codes registered via IANA in rule **MUST** use official HTTP status codes.
- **2021-08-17** : update formats per OpenAPI 3.1 in **MUST** use standard data formats.
- **2021-06-22** : **MUST** use standard data formats changed from **SHOULD** to **MUST**; consistency for rules around standards for data.
- **2021-06-03** : **MUST** secure endpoints with clear distinction of OpenAPI security schemes, favoring `bearer` to `oauth2` .
- **2021-06-01** : resolve uncertainties around 'occurred_at' semantics of event metadata.
- **2021-05-25** : **SHOULD** use standard media types with API endpoint versioning as only custom media type usage exception.
- **2021-05-05** : define usage on resource-ids in `PUT` and `POST` in **MUST** use HTTP methods correctly.
- **2021-04-29** : improve clarity of **MAY** use standard headers.
- **2021-03-19** : clarity on **MUST** use JSON as payload data interchange format.
- **2021-03-15** : **MUST** provide explicit event ordering for data change events changed from **SHOULD** to **MUST**; improve clarity around event ordering.
- **2021-03-19** : best practice section Cursor-based pagination in RESTful APIs
- **2021-02-16** : define how to reference models outside the api in **MUST** only use durable and immutable remote references.
- **2021-02-15** : improve guideline **MUST** support problem JSON — clients must be prepared to not receive problem return objects.
- **2021-01-19** : more details for `GET with body` and `DELETE with body` (**MUST** use HTTP

methods correctly).

- `2020-09-29` : include models for headers to be included by reference in API definitions (**SHOULD** use only the specified proprietary Zalando headers)
- `2020-09-08` : add exception for legacy host names to **MUST** follow naming convention for hostnames
- `2020-08-25` : change **SHOULD** declare enum values using UPPER_SNAKE_CASE string from **MUST** to **SHOULD**, explain exceptions
- `2020-08-25` : add exception for `self` to **MUST** identify resources and sub-resources via path segments and **MUST** pluralize resource names.
- `2020-08-24` : change "**MUST** avoid trailing slashes" to **MUST** use normalized paths without empty path segments and trailing slashes.
- `2020-08-20` : change **SHOULD** use only the specified proprietary Zalando headers from **MUST** to **SHOULD**, mention gateway-specific headers (which are not part of the public API).
- `2020-06-30` : add details to **MUST** use media type versioning
- `2020-05-19` : new sections about DELETE with query parameters and `DELETE with body` in **MUST** use HTTP methods correctly.
- `2020-02-06` : new rule **MAY** expose compound keys as resource identifiers
- `2020-02-05` : add Sunset header, clarify deprecation producedure (**MUST** obtain approval of clients before API shut down, **MUST** collect external partner consent on deprecation time span, **MUST** reflect deprecation in API specifications, **MUST** monitor usage of deprecated API scheduled for sunset, **SHOULD** add `Deprecation` and `Sunset` header to responses, **SHOULD** add monitoring for `Deprecation` and `Sunset` header, **MUST** not start using deprecated APIs)
- `2020-01-21` : new rule **SHOULD** declare enum values using UPPER_SNAKE_CASE string (as MUST, changed later to SHOULD)
- `2020-01-15` : change "Warning" to "Deprecation" header in **SHOULD** add `Deprecation` and `Sunset` header to responses, **SHOULD** add monitoring for `Deprecation` and `Sunset` header.
- `2019-10-10` : remove never-implemented rule "**MUST** Permissions on events must correspond to API permissions"
- `2019-09-10` : remove duplicated rule "**MAY** Standards could be used for Language, Country and Currency", upgrade **MUST** use standard formats for country, language and currency

properties from **MAY** to **SHOULD**.

- `2019-08-29` : new rule **MUST** encode binary data in `base64url` , extend **MUST** use JSON as payload data interchange format pointing to RFC-7493
- `2019-08-29` : new rules **SHOULD** design simple query languages using query parameters, **SHOULD** design complex query languages using JSON
- `2019-07-30` : new rule **MUST** use standard data formats
- `2019-07-30` : change **MUST** use the common money object from **SHOULD** to **MUST**
- `2019-07-30` : change "**SHOULD** Null values should have their fields removed to" **MUST** use same semantics for `null` and absent properties.
- `2019-07-25` : new rule **SHOULD** use naming convention for date/time properties.
- `2019-07-18` : improved cursor guideline for `GET with body` .
- `2019-06-25` : change **MUST** define collection format of header and query parameters from **SHOULD** to **MUST**, use OpenAPI 3 syntax
- `2019-06-13` : remove `X-App-Domain` from **SHOULD** use only the specified proprietary Zalando headers.
- `2019-05-17` : add `X-Mobile-Advertising-Id` to **SHOULD** use only the specified proprietary Zalando headers.
- `2019-04-09` New rule **MUST** only use durable and immutable remote references
- `2019-02-19` : New rule **MUST** support `X-Flow-ID` extracted + expanded from **SHOULD** use only the specified proprietary Zalando headers.
- `2019-01-24:` Improve guidance on caching (**MUST** fulfill common method properties, **MUST** document cacheable `GET` , `HEAD` , and `POST` endpoints).
- `2019-01-21:` Improve guidance on idempotency, introduce idempotency-key (**SHOULD** consider to design `POST` and `PATCH` idempotent, **SHOULD** use secondary key for idempotent `POST` design).
- `2019-01-16` : Change **SHOULD** not use /api as base path from **MAY** to {SHOULD NOT}
- `2018-10-19` : Add `ordering_key_field` to event type definition schema (**MUST** specify and register events as event types, **SHOULD** provide explicit event ordering for general events)
- `2018-09-28` : New rule **MUST** use URL-friendly resource identifiers
- `2018-09-13` : replaced OpenAPI 2.0 syntax with OpenAPI 3.0 in the example snippets

- **`2018-08-10`** : New rule **MUST** document implicit response filtering
- **`2018-07-12`** : Add **`audience`** field to event type definition (**MUST** specify and register events as event types)
- **`2018-06-11:`** Introduced new naming guidelines for host, permission, and event names.
- **`2018-01-10:`** Moved meta information related aspects into new chapter REST Basics - Meta information.
- **`2018-01-09:`** Changed publication requirements for API specifications (**MUST** publish OpenAPI specification for non-component-internal APIs).
- **`2017-12-07:`** Added best practices section including discussion about optimistic locking approaches.
- **`2017-11-28:`** Changed OAuth flow example from password to client credentials in REST Basics - Security.
- **`2017-11-22:`** Updated description of X-Tenant-ID header field
- **`2017-08-22:`** Migration to Asciidoc
- **`2017-07-20:`** Be more precise on client vs. server obligations for compatible API extensions.
- **`2017-06-06:`** Made money object guideline clearer.
- **`2017-05-17:`** Added guideline on query parameter collection format.
- **`2017-05-10:`** Added the convention of using RFC2119 to describe guideline levels, and replaced **`book.could`** with **`book.may`** .
- **`2017-03-30:`** Added rule that permissions on resources in events must correspond to permissions on API resources
- **`2017-03-30:`** Added rule that APIs should be modelled around business processes
- **`2017-02-28:`** Extended information about how to reference sub-resources and the usage of composite identifiers in the **MUST** identify resources and sub-resources via path segments part.
- **`2017-02-22:`** Added guidance for conditional requests with If-Match/If-None-Match
- **`2017-02-02:`** Added guideline for batch and bulk request
- **`2017-02-01:`** **SHOULD** use **`Location`** header instead of **`Content-Location`** header
- **`2017-01-18:`** Removed "Avoid Javascript Keywords" rule

- **`2017-01-05:`** Clarification on the usage of the term "REST/RESTful"
- **`2016-12-07:`** Introduced "API as a Product" principle
- **`2016-12-06:`** New guideline: "Should Only Use UUIDs If Necessary"
- **`2016-12-04:`** Changed OAuth flow example from implicit to password in REST Basics - Security.
- **`2016-10-13:`** **SHOULD** use standard media types
- **`2016-10-10:`** Introduced the changelog. From now on all rule changes on API guidelines will be recorded here.

---

1. Per definition of R.Fielding REST APIs have to support HATEOAS (maturity level 3). Our guidelines do not strongly advocate for full REST compliance, but limited hypermedia usage, e.g. for pagination (see REST Design - Hypermedia). However, we still use the term "RESTful API", due to the absence of an alternative established term and to keep it like the very majority of web service industry that also use the term for their REST approximations — in fact, in today's industry full HATEOAS compliant APIs are a very rare exception.
2. HTTP/1.1 standard (RFC 9110 Section 7.6.1) defines two types of headers: *end-to-end* and *hop-by-hop* headers. End-to-end headers must be transmitted to the ultimate recipient of a request or response. Hop-by-hop headers, on the contrary, are meaningful for a single connection only.

Last updated 2025-05-27 13:43:53 UTC