

Table of Contents

Zalando RESTful API and Event Scheme Guidelines	1
1. Introduction	1
Conventions Used in These Guidelines	2
Zalando specific information	2
2. Principles	2
API as a Product	2
API Design Principles	3
3. General Guidelines	4
Must: Follow API First Principle	4
Must: Provide API Specification using OpenAPI	5
Should: Provide API User Manual	5
Must: Write APIs in U.S. English	5
4. Meta Information	5
Must: Contain API Meta Information	6
Must: Use Semantic Versioning	6
Must: Provide API Audience	7
Should: Provide API Identifiers	8
5. Security	9
Must: Secure Endpoints with OAuth 2.0	9
Must: Define and Assign Access Rights (Scopes)	10
6. Compatibility	11
Must: Don't Break Backward Compatibility	11
Should: Prefer Compatible Extensions	12
Must: Prepare Clients To Not Crash On Compatible API Extensions	12
Should: Design APIs Conservatively	13
Must: Always Return JSON Objects As Top-Level Data Structures To Support Extensibility	14
Must: Treat Open API Definitions As Open For Extension By Default	14
Should: Used Open-Ended List of Values (x-extensible-enum) Instead of Enumerations	14
Should: Avoid Versioning	15
Must: Use Media Type Versioning	15
Must: Do Not Use URI Versioning	16
7. Deprecation	16
Must: Obtain Approval of Clients	16
Must: External Partners Must Agree on Deprecation Timespan	17
Must: Reflect Deprecation in API Definition	17
Must: Monitor Usage of Deprecated APIs	17
Should: Add a Warning Header to Responses	17
Should: Add Monitoring for Warning Header	17

Must: Not Start Using Deprecated APIs	18
8. JSON Guidelines	18
Must: Property names must be ASCII snake_case (and never camelCase): <code>^[a-z_][a-z_0-9]*\$</code> ..	18
Should: Define maps using <code>additionalProperties</code> , without restriction on key format.	18
Should: Array names should be pluralized	19
Must: Boolean property values must not be null.	19
Should: Null values should have their fields removed	20
Should: Empty array values should not be null	20
Should: Enumerations should be represented as Strings	20
Should: Date property values should conform to RFC 3339	20
May: Time durations and intervals could conform to ISO 8601	21
May: Standards could be used for Language, Country and Currency	21
9. API Naming	21
Must: Use lowercase separate words with hyphens for Path Segments	21
Must: Use snake_case (never camelCase) for Query Parameters	21
Should: Prefer Hyphenated-Pascal-Case for HTTP header Fields	22
Must: Pluralize Resource Names	22
May: Use /api as first Path Segment	22
Must: Avoid Trailing Slashes	22
May: Use Conventional Query Strings	22
10. Resources	23
Must: Avoid Actions — Think About Resources	23
Should: Model complete business processes	23
Should: Define <i>useful</i> resources	23
Must: Keep URLs Verb-Free	24
Must: Use Domain-Specific Resource Names	24
Must: Identify resources and Sub-Resources via Path Segments.	24
Should: Only Use UUIDs If Necessary	25
May: Consider Using (Non-) Nested URLs	25
Should: Limit number of Resource types	26
Should: Limit number of Sub-Resource Levels	26
11. HTTP	27
Must: Use HTTP Methods Correctly	27
Must: Fulfill Safeness and Idempotency Properties	30
Must: Use Specific HTTP Status Codes	31
Must: Specify Success and Error Responses	33
Must: Use 207 for Batch or Bulk Requests	34
Must: Use 429 with Headers for Rate Limits	34
Should: Explicitly define the Collection Format of Query Parameters	35
12. Performance	35
Should: Reduce Bandwidth Needs and Improve Responsiveness	35

Should: Use gzip Compression	36
Should: Support Filtering of Resource Fields	36
Should: Allow Optional Embedding of Sub-Resources	38
Must: Document Caching, if Supported	39
13. Pagination	39
Must: Support Pagination	39
Should: Prefer Cursor-Based Pagination, Avoid Offset-Based Pagination	40
May: Use Pagination Links Where Applicable	41
14. Hypermedia	41
Must: Use REST Maturity Level 2	42
May: Use REST Maturity Level 3 - HATEOAS	42
Must: Use full, absolute URI	43
Must: Use Common Hypertext Controls	43
Should: Use Simple Hypertext Controls for Pagination and Self-References	44
Must: Not Use Link Headers with JSON entities	44
15. Data Formats	44
Must: Use JSON to Encode Structured Data	44
May: Use non JSON Media Types for Binary Data or Alternative Content Representations	45
Must: Use Standard Date and Time Formats	45
May: Use Standards for Country, Language and Currency Codes	45
Must: Define Format for Type Number and Integer	45
Should: Prefer standard Media type name application/json	46
16. Common Data Types	46
Should: Use a Common Money Object	47
Must: Use common field names and semantics	48
Must: Follow Hypertext Control Conventions	51
Must: Use Problem JSON	51
Must: Do not expose Stack Traces	52
17. Common Headers	52
Must: Use Content Headers Correctly	52
May: Use Standardized Headers	52
May: Use Content-Location Header	53
Should: Use Location Header instead of Content-Location Header	53
May: Use the Prefer header to indicate processing preferences	53
May: Consider using ETag together with If-(None-)Match header	54
18. Proprietary Headers	55
Must: Use Only the Specified Proprietary Zalando Headers	56
Must: Propagate Proprietary Headers	57
19. API Operation	58
Must: Publish OpenAPI Specification	58
Should: Monitor API Usage	58

20. Events	58
Must: Treat Events as part of the service interface.....	59
Must: Make Event schema available for review	59
Must: Ensure Event schema conforms to Open API Schema Object	59
Must: Ensure that Events are registered as Event Types	60
Must: Ensure Events conform to a well-known Event Category	63
Must: Ensure that Events define useful business resources	67
Must: Events must not provide sensitive customer personal data	67
Must: Use the General Event Category to signal steps and arrival points in business processes ..	68
Must: Use Data Change Events to signal mutations	68
Should: Provide a means for explicit event ordering.....	68
Should: Use the hash partition strategy for Data Change Events	69
Should: Ensure that Data Change Events match API representations	69
Must: Permissions on events must correspond to API permissions	70
Must: Indicate ownership of Event Types	70
Must: Define Event Payloads in accordance with the overall Guidelines	70
Must: Maintain backwards compatibility for Events	71
Should: Avoid additionalProperties in event type definitions	71
Must: Use unique Event identifiers	72
Should: Design for idempotent out-of-order processing	72
Must: Follow conventions for Event Type names	73
Must: Prepare for duplicate Events	73
Appendix A: References	73
OpenAPI Specification	73
Publications, specifications and standards	73
Dissertations	74
Books.....	74
Blogs	74
Appendix B: Tooling.....	74
API First Integrations	74
Support Libraries.....	75
Appendix C: Best Practices	75
Optimistic Locking in RESTful APIs	75
Appendix D: Changelog	79
Rule Changes	80

Zalando RESTful API and Event Scheme Guidelines

Zalando SE



Other formats: [PDF](#), [EPUB3](#)

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 most purely express 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 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 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, self-descriptive messages and hypermedia. RESTful APIs tend to be less use-case specific and comes with less rigid client / server coupling and are more suitable as a platform interface being open for diverse client applications.

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

We apply the RESTful web service principles to all kind of application components, whether they provide functionality via the Internet or via the intranet as larger application elements. We strive to build interoperating distributed systems that different teams can evolve in parallel.

An important principle for (RESTful) API design and usage is Postel's Law, aka [the Robustness Principle](#) (RFC 1122):

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

Readings: Read the following to gain additional insight on the RESTful service architecture paradigm and general RESTful API design style:

- Book: [Irresistable APIs: Designing web APIs that developers will love](#)
- Book: [REST in Practice: Hypermedia and Systems Architecture](#)
- Book: [Build APIs You Won't Hate](#)
- InfoQ eBook: [Web APIs: From Start to Finish](#)

- Lessons-learned blog: [Thoughts on RESTful API Design](#)
- Fielding Dissertation: [Architectural Styles and the Design of Network-Based Software Architectures](#)

3. General Guidelines

The titles are marked with the corresponding labels: **Must:**, **Should:**, **May:**.

Must: Follow API First Principle

As mentioned in the introduction, API First is one of our [engineering and architecture principles](#). In a nutshell API First requires two aspects:

- define APIs outside the code first 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

An element of API First are also this API Guidelines and a [lightweight API review process \[internal link\]](#) 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.

Must: Provide API Specification using OpenAPI

We use the [OpenAPI specification](#) (aka Swagger specification) as standard to define API specifications files. API designers have to provide the API specification files using **YAML** to improve readability.

Please stick to OpenAPI 2.0 for now, until we succeed to update all our tooling to support the upcoming OpenAPI 3.0.0.

The API specification files should be subject to version control using a source code management system - best together with the implementing sources. The API specifications of component external APIs **must be published** and **marked with the intended target audience**.

The [Zalando API Linter Zally](#) should be used to design APIs compliant with this API guideline. Please have also a look to the [API review procedures \[internal link\]](#).

Should: Provide API User Manual

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 GHE pages, on specific team web servers, or as a Google doc. 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 in U.S. English

4. Meta Information

Must: Contain API Meta Information

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

- `#/info/title` as (unique) identifying 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-audience` intended target audience of the API ([see rule 219](#))

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

- `#/info/x-api-id` unique identifier of the API ([see rule 215](#))

Must: Use Semantic Versioning

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 this 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:

```
swagger: '2.0'
info:
  title: Parcel Service API
  description: API for <...>
  version: 1.3.7
<...>
```

Must: Provide API Audience

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

The API consumers with this audience are restricted to applications of the same [functional component](#) ([internal link](#)). All services of a functional component are owned by specific dedicated owner and engineering team. Typical examples 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 Open API 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](#)).

Example:

```
swagger: '2.0'  
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](#).

Should: Provide API Identifiers

Each API should be identified by an explicit, owner assigned, globally unique, and immutable API identifier. APIs evolve and every API aspect may change, except the API identifier. Based on the API identifier, we can track the API life cycle and manage the history and evolution of an API as a sequence of API specifications. For instance, we can provide our [API Linter Zally](#) with the capability to check compatibility and correct semantic versioning of an API update draft against an API running in production.

The API identifier is provided in the **info**-block of the Open API specification and must conform to the following specification:

```
/info/x-api-id:  
  type: string  
  format: urn  
  pattern: ^[a-z0-9][a-z0-9-:.]{6,62}[a-z0-9]$\br/>  description: |  
    Globally unique and immutable ID required to identify the API. The API  
    identifier allows to track the history and evolution of an API as a  
    sequence of API specifications.
```

It is responsibility of the API owner to ensure that the application of the API identifier together with the semantic API [version](#) reflect the expected API history.

While it is nice to use human readable API identifiers based on self-managed URNs, it is recommend to stick to UUIDs to relief API designers from any urge of changing the API identifier while evolving the API. Example:

```
swagger: '2.0'  
info:  
  x-api-id: d0184f38-b98d-11e7-9c56-68f728c1ba70  
  title: Parcel Service API  
  description: API for <...>  
  version: 1.5.8  
  <...>
```

For more information see [API Identifier narrative](#).

5. Security

Must: Secure Endpoints with OAuth 2.0

Every API endpoint needs to be secured using OAuth 2.0. Please refer to the [official OpenAPI spec](#) on how to specify security definitions in you API specification or take a look at the following example.

```
securityDefinitions:  
  oauth2:  
    type: oauth2  
    flow: application  
    tokenUrl: https://identity.zalando.com/oauth2/token  
    scopes:  
      fulfillment-order-service.read: Access right needed to read from the fulfillment  
order service.  
      fulfillment-order-service.write: Access right needed to write to the fulfillment  
order service.
```

The example defines OAuth2 with client credentials flow as security standard used for authentication when accessing endpoints; additionally, there are two API access rights defined via the scopes section for later endpoint authorization usage - please see next section.

It makes little sense specifying the flow to retrieve OAuth tokens in the `securityDefinitions` section, as API endpoints should not care, how OAuth tokens were created. Unfortunately the `flow` field is mandatory and cannot be omitted. API endpoints should always set `flow: application` and ignore this information.

Must: Define and Assign Access Rights (Scopes)

Every API needs to define access rights, called scopes here, and every endpoint needs to have at least one scope assigned. Scopes are defined by name and description per API specification, as shown in the previous section. Please refer to the following rules when creating scope names:

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

<api-standard-scope> ::= <application-id>.<access-type>
<api-resource-specific-scope> ::= <application-id>.<resource-id>.<access-type>
<api-pseudo-scope> ::= uid

<application-id> ::= <as defined via STUPS>
<access-type> ::= read | write -- might be extended in future
<resource-id> ::= <free identifier following application-id syntax>
```

APIs should stick to standard scopes by default — for the majority of use cases, restricting access to specific APIs (with read vs. write differentiation) is sufficient for controlling access for client types like merchant or retailer business partners, customers or operational staff. We want to avoid too many, fine grained scopes increasing governance complexity without real value add. In some situations, where the API serves different types of resources for different owners, resource specific scopes may make sense.

Some examples for standard and resource-specific scopes:

Application ID	Resource ID	Access Type	Example
fulfillment-order		read	fulfillment-order.read
fulfillment-order		write	fulfillment-order.write
sales-order	sales_order	read	sales-order.sales_order.read
sales-order	shipment_order	read	sales-order.shipment_order.read

After scopes names are defined and the scope is declared in the security definition at the top of an API specification, it should be assigned to each API operation by specifying a `security requirement` like this:

```
paths:
  /sales-orders/{order-number}:
    get:
      summary: Retrieves a sales order
      security:
        - oauth2:
            - sales-order-service.sales_order.read
```

In very rare cases a whole API or some selected endpoints may not require specific access control. However, to make this explicit you should assign the `uid` pseudo access right scope in this case. It is the user id and always available as OAuth2 default scope.

```
paths:
  /public-information:
    get:
      summary: Provides public information about ...
                Accessible by any user; no access rights needed.
      security:
        - oauth2:
            - uid
```

Hint: you need not explicitly define the "Authorization" header; it is a standard header so to say implicitly defined via the security section.

6. Compatibility

Must: Don't Break Backward Compatibility

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

We strongly encourage using compatible API extensions and discourage versioning (see rules [113](#), [114](#) below). The following guidelines for service providers ([rule 107](#)) and consumers ([rule 108](#)) 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

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. Enum range can be reduced when used as output parameters.
- Enum ranges cannot not 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.
- Use `x-extensible-enum`, if range is used for output parameters and likely to be extended with growing functionality. It defines an open list of explicit values and clients must be agnostic to new values.
- Support redirection in case an URL has to change ([301 Moved Permanently](#)).

Must: Prepare Clients To Not Crash On Compatible API Extensions

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 for 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 service providers:

- 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 parameter may deliver new values; either be agnostic or provide default behavior for unknown values.

- 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 x00 code (see [RFC7231 Section 6](#)).
- Follow the redirect when the server returns HTTP status [301 Moved Permanently](#).

Should: Design APIs Conservatively

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.
- 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 [RFC7231 Section 4.3.4](#)). 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: Always Return JSON Objects As Top-Level Data Structures To Support Extensibility

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.

Must: Treat Open API Definitions As Open For Extension By Default

The Open API 2.0 specification is not very specific on default extensibility of objects, and redefines JSON-Schema keywords related to extensibility, like `additionalProperties`. Following our overall compatibility guidelines, Open API object definitions are considered open for extension by default as per [Section 5.18 "additionalProperties"](#) of JSON-Schema.

When it comes to Open API 2.0, 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. Instead of ignoring fields, servers *may* 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. API designers must document clearly how unexpected fields are 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.

Should: Used Open-Ended List of Values (x-extensible-enum) Instead of Enumerations

Enumerations are per definition closed sets of values, that are 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 strongly 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 value is complete with respect to any thinkable and unthinkable future feature.

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

```
deliver_methods:
  type: string
  x-extensible-enum:
    - parcel
    - letter
    - email
```

Note: `x-extensible-enum` is not JSON Schema conform but will be ignored by most tools.

Should: Avoid Versioning

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

However, when API versioning is unavoidable, you have to design your multi-version RESTful APIs using media type versioning (instead of URI versioning, see below). Media type versioning is less tightly coupled since it supports content negotiation and hence reduces complexity of release management.

Media type versioning: Here, 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. Note: This versioning only applies to the request and response content schema, not to URI or method semantics.

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
```

Using header versioning should:

- include versions in request and response headers to increase visibility
- include Content-Type in the Vary header to enable proxy caches to differ between versions

Hint: Until an incompatible change is necessary, it is recommended to stay with the standard `application/json` media type.

Hint: [OpenAPI currently doesn't support content negotiation](#), though [a comment in this issue](#) mentions a workaround (using a fragment identifier that gets stripped off). Another way would be to document just the new version, but let the server accept the old one (with the previous content-type).

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

Must: Do Not Use URI Versioning

With URI 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 URI versioning, and go instead with media type versioning and content negotiation (see above).

7. Deprecation

Sometimes it is necessary to phase out an API endpoint (or version), for instance, if a field is no longer supported in the result or a whole business functionality behind an endpoint has to be shut down. There are many other reasons as well. As long as these endpoints are still used by consumers these are breaking changes and not allowed. Deprecation rules have to be applied to make sure that necessary consumer changes are aligned and deprecated endpoints are not used before API changes are deployed.

Must: Obtain Approval of Clients

Before shutting down an API (or version of an API) the producer must make sure, that all clients have given their consent to shut down the endpoint. Producers should help consumers to migrate to a potential new endpoint (i.e. by providing a migration manual). After all clients are migrated, the producer may shut down the deprecated API.

Must: External Partners Must Agree on Deprecation Timespan

If the API is consumed by any external partner, the producer must define a reasonable timespan that the API will be maintained after the producer has announced deprecation. The external partner (client) must agree to this minimum after-deprecation-lifespan before he starts using the API.

Must: Reflect Deprecation in API Definition

API deprecation must be part of the OpenAPI definition. If a method on a path, a whole path or even a whole API endpoint (multiple paths) should be deprecated, the producers must set `deprecated=true` on each method / path element that will be deprecated (OpenAPI 2.0 only allows you to define deprecation on this level). If deprecation should happen on a more fine grained level (i.e. query parameter, payload etc.), the producer should set `deprecated=true` on the affected method / path element and add further explanation to the `description` section.

If `deprecated` is set to `true`, the producer must describe what clients should use instead and when the API will be shut down in the `description` section of the API definition.

Must: Monitor Usage of Deprecated APIs

Owners of APIs used in production must monitor usage of deprecated APIs until the API can be shut down in order to align deprecation and avoid uncontrolled breaking effects. See also the [Should: Monitor API Usage](#).

Should: Add a Warning Header to Responses

During deprecation phase, the producer should add a `Warning` header (see [RFC 7234 - Warning header](#)) field. When adding the `Warning` header, the `warn-code` must be `299` and the `warn-text` should be in form of *"The path/operation/parameter/... {name} is deprecated and will be removed by {date}. Please see {link} for details."* with a link to a documentation describing why the API is no longer supported in the current form and what clients should do about it. Adding the `Warning` header is not sufficient to gain client consent to shut down an API.

Should: Add Monitoring for Warning Header

Clients should monitor the `Warning` header in HTTP responses to see if an API will be deprecated in future.

Must: Not Start Using Deprecated APIs

Clients must not start using deprecated parts of an API.

8. JSON Guidelines

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.

The first some of the following guidelines are about property names, the later ones about values.

Must: Property names must be ASCII snake_case (and never camelCase): `^[a-z_][a-z_0-9]*$`

Property names are restricted to ASCII strings. The first character must be a letter, or an underscore, and subsequent characters can be a letter, an underscore, or a number.

(It is recommended to use `_` at the start of property names only for keywords like `_links`.)

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: Define maps using `additionalProperties`, without restriction on key format.

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 for 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):

```

definitions:
  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 https://tools.ietf.org/html/bcp47[BCP-47 language tags].
        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"
  }
}

```

Should: Array names should be pluralized

To indicate they contain multiple values prefer to pluralize array names. This implies that object names should in turn be singular.

Must: Boolean property values must not be null

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, strongly prefer to replace the boolean with enumeration of named values or statuses - for example `accepted_terms_and_conditions` with true or false can be replaced with `terms_and_conditions` with values yes, no and unknown.

Should: Null values should have their fields removed

Swagger/OpenAPI, which is in common use, doesn't support null field values (it does allow omitting that field completely if it is not marked as required). However that doesn't prevent clients and servers sending and receiving those fields with null values. Also, in some cases null may be a meaningful value - for example, JSON Merge Patch [RFC 7382](#)) using null to indicate property deletion.

Should: Empty array values should not be null

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

Should: Enumerations should be represented as Strings

Strings are a reasonable target for values that are by design enumerations.

Should: Date property values should conform to RFC 3339

Use the date and time formats defined by [RFC 3339](#):

- for "date" use strings matching `date-fullyear` "-" `date-month` "-" `date-mday`, for example: `2015-05-28`
- for "date-time" use strings matching `full-date` "T" `full-time`, for example `2015-05-28T14:07:17Z`

Note that the [OpenAPI format](#) "date-time" corresponds to "date-time" in the RFC) and `2015-05-28` for a date (note that the OpenAPI format "date" corresponds to "full-date" in the RFC). Both are specific profiles, a subset of the international standard [ISO 8601](#).

A zone offset may be used (both, in request and responses)—this is simply defined by the standards. However, we encourage restricting dates to UTC and without 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 that might be including daylight saving time. Localization of dates should be done by the services that provide user interfaces, if required.

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.

Sometimes it can seem data is naturally represented using numerical timestamps, but this can introduce interpretation issues with precision - for example 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.

May: Time durations and intervals could conform to ISO 8601

Schema based JSON properties that are by design durations and intervals could be strings formatted as recommended by ISO 8601 ([Appendix A of RFC 3339 contains a grammar](#) for durations).

May: Standards could be used for Language, Country and Currency

- [ISO 3166-1-alpha2 country](#)
- (It's "GB", not "UK", even though "UK" has seen some use at Zalando)
- [ISO 639-1 language code](#)
- [BCP-47](#) (based on ISO 639-1) for language variants
- [ISO 4217 currency codes](#)

9. API Naming

Must: Use lowercase separate words with hyphens for Path Segments

Example:

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

This applies to concrete path segments and not the names of path parameters. For example `{shipment_order_id}` would be ok as a path parameter.

Must: Use snake_case (never camelCase) for Query Parameters

Examples:

```
customer_number, order_id, billing_address
```

Should: Prefer Hyphenated-Pascal-Case for HTTP header Fields

This is for consistency in your documentation (most other headers follow this convention). Avoid camelCase (without hyphens). Exceptions are common abbreviations like “ID.”

Examples:

```
Accept-Encoding  
Apply-To-Redirect-Ref  
Disposition-Notification-Options  
Original-Message-ID
```

See also: [HTTP Headers are case-insensitive \(RFC 7230\)](#).

See [Common Headers](#) and [Proprietary Headers](#) sections for more guidance on HTTP headers.

Must: Pluralize Resource Names

Usually, a collection of resource instances is provided (at least API should be ready here). The special case of a resource singleton is a collection with cardinality 1.

May: Use /api as first Path Segment

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 — add “/api” as base path to clearly separate public and non-public API resources.

Must: Avoid Trailing Slashes

The trailing slash must not have specific semantics. Resource paths must deliver the same results whether they have the trailing slash or not.

May: Use Conventional Query Strings

If you provide query support for sorting, pagination, filtering functions or other actions, use the following standardized naming conventions:

- **q** — default query parameter (e.g. used by browser tab completion); should have an entity specific alias, like sku
- **limit** — to restrict the number of entries. See [Pagination](#) section below. Hint: You can use size as an alternate query string.
- **cursor** — key-based page start. See [Pagination](#) section below.

- **offset** — numeric offset page start. See [Pagination](#) section below. Hint: In combination with limit, you can use page as an alternative to offset.
- **sort** — comma-separated list of fields to sort. To indicate sorting direction, fields may be prefixed with + (ascending) or - (descending, default), e.g. /sales-orders?sort=+id
- **fields** — to retrieve a subset of fields. See [Support Filtering of Resource Fields](#) below.
- **embed** — to expand embedded entities (ie.: inside of an article entity, expand silhouette code into the silhouette object). Implementing “expand” correctly is difficult, so do it with care.

10. Resources

Must: Avoid Actions — Think About Resources

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: Model complete business processes

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.

Should: Define *useful* resources

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: Keep URLs Verb-Free

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: Use Domain-Specific Resource Names

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.

Must: Identify resources and Sub-Resources via Path Segments

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.

Composite identifiers must not contain `/` as a separator. 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 example, if `/customers/12ev123bv12v/addresses/DE_100100101` is a valid path of your API, then `/customers/12ev123bv12v/addresses`, `/customers/12ev123bv12v` and `/customers` must be valid as well in principle.

Basic URL structure:

```
{resources}/{resource-id}/{sub-resources}/{sub-resource-id}
{resources}/{partial-id-1}[separator][partial-id-2]
```

Examples:

```
/carts/1681e6b88ec1/items
/carts/1681e6b88ec1/items/1
/customers/12ev123bv12v/addresses/DE_100100101
```

Should: Only Use UUIDs If Necessary

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 roundtrips.

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.

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.

May: Consider Using (Non-) Nested URLs

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:

```
/carts/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/1681e6b88ec1  
/sales-orders/5273gh3k525a
```

Should: Limit number of Resource types

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

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)

11. HTTP

Must: Use HTTP Methods Correctly

Be compliant with the standardized HTTP method semantics summarized as follows:

GET

GET requests are used to read a single resource or query set of resources.

- 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 listing is empty) or 404 (if the list is missing)
- GET requests must NOT have request body payload

Note: GET requests on collection resources should provide a sufficient filter mechanism as well as [Pagination](#).

"GET with Body"

APIs sometimes face the problem, that they have to provide extensive structured request information with GET, that may even conflicts with the size limits of clients, load-balancers, and servers. As we require APIs to be standard conform (body 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 distributed to multiple query parameters or a single structured URL encoded string.
2. POST with body content: when a GET with URL encoded query parameters is not possible, a POST with body content must be used. In this case the endpoint must be documented with the hint **GET with body** to transport the GET semantic of this call.

Note: It is no option to encode the lengthy structured request information in header parameters. From a conceptual point of view, the semantic of an operation should always be expressed by resource name and query parameters, i.e. what goes into the URL. Request headers are reserved for general context information, e.g. FlowIDs. 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.

PUT

PUT requests are used to create or update **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 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)
- successful PUT requests will usually generate 200 or 204 (if the resource was updated - with or without actual content returned), and 201 (if the resource was created)

Note: Resource IDs with respect to PUT requests are maintained by the client and passed as a URL path segment. Putting the same resource twice is required to be idempotent and to result in the same single resource instance. If PUT is applied for creating a resource, only URIs should be allowed as resource IDs. If URIs are not available POST should be preferred.

To prevent unnoticed concurrent updates when using PUT, the combination of **ETag** and **If-(None-)Match** headers should be considered to signal the server stricter demands to expose conflicts and prevent lost updates. The section [Optimistic Locking in RESTful APIs](#) also describes some alternatives to this approach.

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 collection resource identified by the URL«*.

- POST request should only be applied to collection resources, and normally not on single resource, as this has an undefined semantic
- on successful POST requests, the server will create one or multiple new resources and provide their URI/URLs in the response
- successful POST requests will usually generate 200 (if resources have been updated), 201 (if resources have been created), and 202 (if the request was accepted but has not been finished yet)

More generally: POST should be used for scenarios that cannot be covered by the other methods sufficiently. For instance, GET with complex (e.g. SQL like structured) query that needs to be passed as request body payload because of the URL-length constraint. In such cases, make sure to document the fact that POST is used as a workaround.

Note: Resource IDs with respect to POST requests are created and maintained by server and returned with response payload. Posting the same resource twice is by itself **not** required to be idempotent and may result in multiple resource instances. Anyhow, if external URIs are present that can be used to identify duplicate requests, it is best practice to implement POST in an idempotent way.

PATCH

PATCH request are only used for partial update of single resources, i.e. where only a specific subset of resource fields should be replaced. The semantic is best described as *»please change the resource identified by the URL according to my change request«*. The semantic of the change request is not defined in the HTTP standard and must be described in the API specification by using suitable media types.

- PATCH requests are usually applied to single resources, and not on collection resources, as this would imply patching on the entire collection
- 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 will usually generate 200 or 204 (if resources have been updated)
- with or without updated content returned)

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).
2. use PATCH with partial objects to only update parts of a resource, whenever possible. (This is basically [JSON Merge Patch](#), a specialized media type `application/merge-patch+json` that is a partial resource representation.)
3. use PATCH with [JSON Patch](#), 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 media type.

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 cases [JSON Patch](#) can shown its full power while still showing readable patch requests (see also [JSON patch vs. merge](#)).

To prevent unnoticed concurrent updates when using PATCH, the combination of `ETag`` and ``If-Match` headers should be considered to signal the server stricter demands to expose conflicts and prevent lost updates.

DELETE

DELETE request 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
- successful DELETE request will usually generate 200 (if the deleted resource is returned) or 204

(if no content is returned)

- failed DELETE request will usually generate 404 (if the resource cannot be found) or 410 (if the resource was already deleted before)

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.

OPTIONS

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

- OPTIONS requests usually either return a comma separated list of methods (provided by an **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 Safeness and Idempotency Properties

An operation can be...

- idempotent, i.e. operation will have the same effect on the server's state if executed once or multiple times (note: this does not necessarily mean returning the same response or status code)
- safe, i.e. must not have side effects such as state changes

Method implementations must fulfill the following basic properties:

HTTP method	safe	idempotent
OPTIONS	Yes	Yes
HEAD	Yes	Yes
GET	Yes	Yes
PUT	No	Yes
POST	No	No
DELETE	No	Yes
PATCH	No	No

Please see also [Best Practices \[internal link\]](#) for further hints on how to support the different HTTP methods on resources.

Must: Use Specific HTTP Status Codes

This guideline groups the following rules for HTTP status codes usage:

- You must not invent new HTTP status codes; only use standardized HTTP status codes and consistent with its intended semantics.
- You should use the most specific HTTP status code for your concrete resource request processing status or error situation.
- When using HTTP status codes that are less commonly used and not listed below, you must provide good documentation in the API definition.

There are ~60 different HTTP status codes with specific semantics defined in the HTTP standards (mainly [RFC7231](#) and [RFC-6585](#)) - and there are upcoming new ones, e.g. [draft legally-restricted-status](#) (see overview on all error codes on [Wikipedia](#) or via <https://httpstatuses.com/>). And there are unofficial ones, e.g. used by specific web servers like Nginx.

List of most commonly used and best understood HTTP status codes:

Success Codes

Code	Meaning	Methods
200	OK - this is the standard success response	All
201	Created - Returned on successful entity creation. You are free to return either an empty response or the created resource in conjunction with the Location header. (More details found in the Common Headers .) Always set the Location header.	POST, PUT
202	Accepted - The request was successful and will be processed asynchronously.	POST, PUT, DELETE, PATCH
204	No content - There is no response body	PUT, DELETE, PATCH
207	Multi-Status - The response body contains multiple status informations for different parts of a batch/bulk request. See Must: Use 207 for Batch or Bulk Requests .	POST

Redirection Codes

Code	Meaning	Methods
301	Moved Permanently - This and all future requests should be directed to the given URI.	All
303	See Other - The response to the request can be found under another URI using a GET method.	PATCH, POST, PUT, DELETE
304	Not Modified - resource has not been modified since the date or version passed via request headers If-Modified-Since or If-None-Match.	GET

Client Side Error Codes

Code	Meaning	Methods
400	Bad request - generic / unknown error	All
401	Unauthorized - the users must log in (this often means “Unauthenticated”)	All
403	Forbidden - the user is not authorized to use this resource	All
404	Not found - the resource is not found	All
405	Method Not Allowed - the method is not supported, see OPTIONS	All
406	Not Acceptable - resource can only generate content not acceptable according to the Accept headers sent in the request	All
408	Request timeout - the server times out waiting for the resource	All
409	Conflict - request cannot be completed due to conflict, e.g. when two clients try to create the same resource or if there are concurrent, conflicting updates	POST, PUT, DELETE, PATCH
410	Gone - resource does not exist any longer, e.g. when accessing a resource that has intentionally been deleted	All
412	Precondition Failed - returned for conditional requests, e.g. If-Match if the condition failed. Used for optimistic locking.	PUT, DELETE, PATCH
415	Unsupported Media Type - e.g. clients sends request body without content type	POST, PUT, DELETE, PATCH
423	Locked - Pessimistic locking, e.g. processing states	PUT, DELETE, PATCH
428	Precondition Required - server requires the request to be conditional (e.g. to make sure that the “lost update problem” is avoided).	All
429	Too many requests - the client does not consider rate limiting and sent too many requests. See Must: Use 429 with Headers for Rate Limits .	All

Server Side Error Codes:

Code	Meaning	Methods
500	Internal Server Error - a generic error indication for an unexpected server execution problem (here, client retry may be sensible)	All
501	Not Implemented - server cannot fulfill the request (usually implies future availability, e.g. new feature).	All
503	Service Unavailable - server is (temporarily) not available (e.g. due to overload) — client retry may be sensible.	All

Must: Specify Success and Error Responses

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, the OpenAPI specification must contain definitions for **all** success and error responses. Both are part of the interface definition and provide important information for service clients to handle standard as well as exceptional situations. For each response the most specific HTTP status code should be selected that conveys the domain-specific success or error semantic — see [Must: Use Specific HTTP Status Codes](#).

To transport error information beside the status code, we use a standardized error return object specification that must be used — see [Must: Use Problem JSON](#).

Remark: In most cases it is not useful to document all technical errors, especially if they are not under control of the service provider. Thus unless a response code conveys application-specific functional semantics or is used in a none standard way that requires additional explanation, multiple error response specifications can be combined using the following pattern:

```
default:
  description: a specific error occurred - see status code for more information.
  schema:
    $ref: 'https://zalando.github.io/problem/schema.yaml#/Problem'
```

Note: It is best practice to explicitly specify all successful status codes. This prevents any conflict with the above pattern.

The list of status code that in general can be omitted from API specifications includes but is not limited to:

- 401 Unauthorized
- 403 Forbidden
- 404 Not Found unless it has some additional semantics
- 405 Method Not Allowed
- 406 Not Acceptable
- 408 Request Timeout
- 413 Payload Too Large
- 414 URI Too Long
- 415 Unsupported Media Type
- 500 Internal Server Error
- 502 Bad Gateway
- 503 Service Unavailable
- 504 Gateway Timeout

Note: Even though all status code may be documented explicitly, clients must always be prepared for responses with unexpected status codes. In case of doubt they should handle them like they

would handle the corresponding x00 code. Adding new response codes (specially error responses) should be considered a compatible API evolution.

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.

Must: Use 207 for Batch or Bulk Requests

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 an 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** has to respond with HTTP status code **207**, unless it encounters a generic or unexpected failure before looking at individual parts.
- A batch or bulk response with status code 207 **always** returns a multi-status object containing sufficient status and/or monitoring information for each part of the batch or bulk request.
- A batch or bulk request may result in a status code 400/500, only if the service encounters a failure before looking at individual parts or, if an unanticipated failure occurs.

The before rules apply *even in the case* that processing of all individual part *fail* or each part is executed *asynchronously*! They are intended to allow clients to act on batch and bulk responses by inspecting the individual results in a consistent way.

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 429 with Headers for Rate Limits

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 and therefore the request can't be fulfilled. 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.

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 minimise 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.

Should: Explicitly define the Collection Format of Query Parameters

There are different ways of supplying a set of values as a query parameter. One particular type should be selected and stated explicitly in the API definition. The OpenAPI property **collectionFormat** is used to specify the format of the query parameter.

Only the **csv** or **multi** formats should be used for multi-value query parameters as described below.

Collection Format	Description	Example
csv	Comma separated values	?parameter=value1,value2,value3
multi	Multiple parameter instances	?parameter=value1¶meter=value2¶meter=value3

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

12. Performance

Should: Reduce Bandwidth Needs and Improve Responsiveness

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:

- [gzip compression](#)
- querying field filters to retrieve a subset of resource attributes (see [Should: Support Filtering of Resource Fields](#) below)
- paginate lists of data items (see [Pagination](#) below)
- [ETag](#) and [If-\(None-\)Match](#) headers to avoid re-fetching of unchanged resources (see [May: Consider using ETag together with If-\(None-\)Match header](#))
- pagination for incremental access of larger (result) lists

Each of these items is described in greater detail below.

Should: Use gzip Compression

Compress the payload of your API's responses with gzip, unless there's a good reason not to — for example, you are serving so many requests that the time to compress becomes a bottleneck. This helps to transport data faster over the network (fewer bytes) and makes frontends respond faster.

Though gzip compression might be the default choice for server payload, the server should also support payload without compression and its client control via `Accept-Encoding` request header — see also [RFC 7231 Section 5.3.4](#). The server should indicate used gzip compression via the `Content-Encoding` header.

Should: Support Filtering of Resource Fields

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 determine the subset of fields he wants to receive via the `fields` query parameter — example see [Google AppEngine API's partial response](#):

Unfiltered

```
GET http://api.example.org/resources/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",
  "partner": {
    "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/resources/123?fields=(name,partner(name)) HTTP/1.1
```

```
HTTP/1.1 200 OK
```

```
Content-Type: application/json
```

```
{
  "name": "John Doe",
  "partner": {
    "name": "Jane Doe"
  }
}
```

As illustrated by this example, field filtering should be done via request parameter "fields" with value range defined by the following [BNF](#) grammar.

```

<fields> ::= <negation> <fields_expression> | <fields_expression>

<negation> ::= "!"

<fields_expression> ::= "(" <field_set> ")"

<field_set> ::= <qualified_field> | <qualified_field> "," <field_set>

<qualified_field> ::= <field> | <field> <fields_expression>

<field> ::= <DASH_LETTER_DIGIT> | <DASH_LETTER_DIGIT> <field>

<DASH_LETTER_DIGIT> ::= <DASH> | <LETTER> | <DIGIT>

<DASH> ::= "-" | "_"

<LETTER> ::= "A" | "B" | "C" | "D" | "E" | "F" | "G" | "H" | "I" | "J" | "K" | "L" |
"M" | "N" | "O" | "P" | "Q" | "R" | "S" | "T" | "U" | "V" | "W" | "X" | "Y" | "Z" |
"a" | "b" | "c" | "d" | "e" | "f" | "g" | "h" | "i" | "j" | "k" | "l" | "m" | "n" |
"o" | "p" | "q" | "r" | "s" | "t" | "u" | "v" | "w" | "x" | "y" | "z"

<DIGIT> ::= "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"

```

A **fields_expression** as defined by the grammar describes the properties of an object, i.e. **(name)** returns only the **name** property of the root object. **(name,partner(name))** returns the **name** and **partner** properties where **partner** itself is also an object and only its **name** property is returned.

Hint: OpenAPI doesn't allow you to formally specify whether depending on a given parameter will return different parts of the specified result schema. Explain this in English in the parameter description.

Should: Allow Optional Embedding of Sub-Resources

Embedding related resources (also known 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 [May: Use Conventional Query Strings](#) 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 Caching, if Supported

If an API is intended to support caching, it must take care to specify this ability by defining the caching boundaries i.e. time-to-live and cache constraints, by providing the **Cache-Control** and **Vary** headers (Please read [RFC-7234](#) carefully).

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. As a consequence, caching is in best case complex, in worst case inefficient. Thus, client side and transparent HTTP caching should be avoided for RESTful APIs unless the API designer has proven to know better.

As default, API providers should set the **Cache-Control: no-cache** header.

Note: There is no need to document this default setting, that is attached to each response automatically by most frameworks. However, any use deviating from this default must be sufficiently documented.

13. Pagination

Must: Support Pagination

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

There are two page iteration techniques:

- [Offset/Limit-based pagination](#): numeric offset identifies the first page entry

- [Cursor-based](#) — aka key-based — pagination: a unique key element identifies the first page entry (see also [Facebook's guide](#))

The technical conception of pagination should also consider user experience related issues. As mentioned in this [article](#), jumping to a specific page is far less used than navigation via next/previous page links. This favours cursor-based over offset-based pagination.

Should: Prefer Cursor-Based Pagination, Avoid Offset-Based Pagination

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 / limit based pagination is more 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 is not feasible
- Variability of data 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.
 - When using cursor-based pagination, paging cannot continue when the cursor entry has been deleted while fetching two pages
- Performance considerations - efficient server-side processing using offset-based pagination is hardly feasible for:
 - Higher data list volumes, especially if they do not reside in the database's main memory
 - Sharded or NoSQL databases
- Cursor-based navigation may not work if you need the total count of results and / or backward iteration support

Further reading:

- [Twitter](#)
- [Use the Index, Luke](#)
- [Paging in PostgreSQL](#)

May: Use Pagination Links Where Applicable

- API implementing [HATEOS](#) may use [simplified hypertext controls](#) for pagination within collections.

Those collections should then have an `items` attribute holding the items of the current page. The collection may contain additional metadata about the collection or the current page (e.g. `index`, `page_size`) when necessary.

You should avoid providing a total count in your API unless there's a clear need to do so. Very often, there are systems and performance implications to supporting full counts, especially as datasets grow and requests become complex queries or filters that drive full scans (e.g., your database might need to look at all candidate items to count them). While this is an implementation detail relative to the API, it's important to consider your ability to support serving counts over the life of a service.

If the collection consists of links to other resources, the collection name should use [IANA registered link relations](#) as names whenever appropriate, but use plural form.

E.g. a service for articles could represent the collection of hyperlinks to an article's `authors` like that:

```
{
  "self": "https://.../articles/xyz/authors/",
  "index": 0,
  "page_size": 5,
  "items": [
    {
      "href": "https://...",
      "id": "123e4567-e89b-12d3-a456-426655440000",
      "name": "Kent Beck"
    },
    {
      "href": "https://...",
      "id": "987e2343-e89b-12d3-a456-426655440000",
      "name": "Mike Beedle"
    },
    ...
  ],
  "first": "https://...",
  "next": "https://...",
  "prev": "https://...",
  "last": "https://..."
}
```

14. Hypermedia

Must: Use REST Maturity Level 2

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](#)
- [Must: Use Specific 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

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):

- 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 full, absolute URI

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: Use Common Hypertext Controls

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:

```
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**):


```
{
  "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

Hypertext controls for pagination inside collections and self-references should use a simple URI value in combination with their corresponding [link relations](#) ([next](#), [prev](#), [first](#), [last](#), [self](#)) instead of the extensible common hypertext control

See [Pagination](#) for information how to best represent paginateable collections.

Must: Not Use Link Headers with JSON entities

We don't allow the use of the [Link Header defined by RFC 5988](#) in conjunction with JSON media types. We prefer links directly embedded in JSON payloads to the uncommon link header syntax.

15. Data Formats

Must: Use JSON to Encode Structured Data

Use JSON-encoded body payload for transferring structured data. The JSON payload must follow [RFC-7159](#) by having (if possible) a serialized object as the top-level structure, since it would allow for future extension. This also applies for collection resources where one naturally would assume an array. See [May: Use Pagination Links Where Applicable](#) for an example.

May: Use non JSON Media Types for Binary Data or Alternative Content Representations

Other media types may be used in following cases:

- Transferring binary data or data whose structure is not relevant. This is the case if payload structure is not interpreted and consumed by clients as is. Example of such use case is downloading images in formats JPG, PNG, GIF.
- In addition to JSON version alternative data representations (e.g. in formats PDF, DOC, XML) may be made available through content negotiation.

Must: Use Standard Date and Time Formats

JSON Payload

Read more about date and time format in [Should: Date property values should conform to RFC 3339](#).

HTTP headers

Http headers including the proprietary headers use the [HTTP date format defined in RFC 7231](#).

May: Use Standards for Country, Language and Currency Codes

Use the following standard formats for country, language and currency codes:

- [ISO 3166-1-alpha2 country codes](#)
 - (It is “GB”, not “UK”, even though “UK” has seen some use at Zalando)
- [ISO 639-1 language code](#)
 - [BCP-47](#) (based on ISO 639-1) for language variants
- [ISO 4217 currency codes](#)

Must: Define Format for Type Number and Integer

Whenever an API defines a property of type **number** or **integer**, the precision must be defined by the format as follows to prevent clients from guessing the precision incorrectly, and thereby changing the value unintentionally:

type	format	specified value range
integer	int32	integer between -2^{31} and $2^{31}-1$
integer	int64	integer between -2^{63} and $2^{63}-1$
integer	bigint	arbitrarily large signed integer number

type	format	specified value range
number	float	IEEE 754-2008/ISO 60559:2011 binary64 decimal number
number	double	IEEE 754-2008/ISO 60559:2011 binary128 decimal number
number	decimal	arbitrarily precise signed decimal number

The precision must be translated by clients and servers into the most specific language types. E.g. for the following definitions the most specific language types in Java will translate to `BigDecimal` for `Money.amount` and `int` or `Integer` for the `OrderList.page_size`:

```
Money:
  type: object
  properties:
    amount:
      type: number
      description: Amount expressed as a decimal number of major currency units
      format: decimal
      example: 99.95
    ...

OrderList:
  type: object
  properties:
    page_size:
      type: integer
      description: Number of orders in list
      format: int32
      example: 42
```

Should: Prefer standard Media type name `application/json`

Previously, this guideline allowed the use of custom media types like `application/x.zalando.article+json`. This usage is not recommended anymore and should be avoided, except where it is necessary for cases of [media type versioning](#). Instead, the standard media type name `application/json` (or [Must: Use Problem JSON](#)).

Custom media types with subtypes beginning with `x` bring no advantage compared to the standard media type for JSON, and make automated processing more difficult. They are also [discouraged by RFC 6838](#).

16. Common Data Types

Definitions of data objects that are good candidates for wider usage:

Should: Use a Common Money Object

Use the following common money structure:

```
Money:
  type: object
  properties:
    amount:
      type: number
      description: Amount expressed as a decimal number of major currency units
      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
```

The decimal values for "amount" describe unit and subunit of the currency in a single value, where the digits before the decimal point are for the major unit and the digits after the decimal point are for the minor unit. Note 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 \(internal link\)](#), 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.

Must: Use common field names and semantics

There exist a variety of field types that are required in multiple places. To achieve consistency across all API implementations, you must use common field names and semantics whenever applicable.

Generic Fields

There are some data fields that come up again and again in API data:

- **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. **customer_id** not **customer_number**; **parent_node_id** for the reference to a parent node from a child node, even if both have the type **Node**)
- **created**: when the object was created. If used, this must be a **date-time** construct.
- **modified**: when the object was updated. If used, this must be a **date-time** construct.
- **type**: the kind of thing this object is. If used, the type of this field should be a string. Types allow runtime information on the entity provided that otherwise requires examining the Open API file.

Example JSON schema:

```

tree_node:
  type: object
  properties:
    id:
      description: the identifier of this node
      type: string
    created:
      description: when got this node created
      type: string
      format: 'date-time'
    modified:
      description: when got this node last updated
      type: string
      format: 'date-time'
    type:
      type: string
      enum: [ 'LEAF', 'NODE' ]
    parent_node_id:
      description: the identifier of the parent node of this node
      type: string
  example:
    id: '123435'
    created: '2017-04-12T23:20:50.52Z'
    modified: '2017-04-12T23:20:50.52Z'
    type: 'LEAF'
    parent_node_id: '534321'

```

These properties are not always strictly necessary, but making them idiomatic allows API client developers to build up a common understanding of Zalando's resources. There is very little utility for API consumers in having different names or value types for these fields across APIs.

Address Fields

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

```

addressee:
  description: a (natural or legal) person that gets addressed
  type: object
  required:
    - first_name
    - last_name
    - street
    - city
    - zip
    - country_code
  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: Follow Hypertext Control Conventions

APIs that provide hypertext controls (links) to interconnect API resources must follow the conventions for naming and modeling of hypertext controls as defined in section [Hypermedia](#).

Must: Use Problem JSON

[RFC 7807](#) defines the media type `application/problem+json`. Operations should return that (together with a suitable status code) when any problem occurred during processing and you can give more details than the status code itself can supply, whether it be caused by the client or the server (i.e. both for 4xx or 5xx errors).

A previous version of this guideline (before the publication of that RFC and the registration of the media type) told to return `application/x.problem+json` in these cases (with the same contents). Servers for APIs defined before this change should pay attention to the `Accept` header sent by the client and set the `Content-Type` header of the problem response correspondingly. Clients of such APIs should accept both media types.

APIs may define custom problems types with extension properties, according to their specific needs.

The Open API schema definition can be found [on github](#). You can reference it by using:


```
responses:
  503:
    description: Service Unavailable
    schema:
      $ref: 'https://zalando.github.io/problem/schema.yaml#/Problem'
```

Must: Do not expose Stack Traces

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.

17. Common Headers

This section describes a handful of headers, which we found raised the most questions in our daily usage, or which are useful in particular circumstances but not widely known.

Must: Use Content Headers Correctly

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.

May: Use Standardized Headers

Use [this list](#) and mention its support in your OpenAPI definition.

May: Use Content-Location Header

The Content-Location header is *optional* and can be used in successful write operations (PUT, POST or PATCH) or 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 URI 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 URI, 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
```

Should: Use Location Header instead of Content-Location Header

As the correct usage of Content-Location with respect to semantics and caching is difficult, we *discourage* the use of Content-Location. In most cases it is sufficient to direct clients to the resource location by using the Location header instead without hitting the Content-Location specific ambiguities and complexities.

More details in RFC 7231 [7.1.2 Location](#), [3.1.4.2 Content-Location](#)

May: Use the Prefer header to indicate processing preferences

The **Prefer** header defined in [RFC7240](#) allows clients to request processing behaviors from servers. [RFC7240](#) 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 like this in an API definition:

```
Prefer:
  name: Prefer
  description: |
    The RFC7240 Prefer header indicates that particular server
    behaviors are preferred by the client but are not required
    for successful completion of the request.
    # (indicate the preferences supported by the API)

  in: header
  type: string
  required: false
```

Supporting APIs may return the **Preference-Applied** header also defined in [RFC7240](#) to indicate whether the preference was applied.

May: Consider using ETag together with If-(None-)Match header

When creating or updating resources it may be necessary to expose conflicts and to prevent the lost update problem. This can be best accomplished by using the **ETag** header together with the **If-Match** and **If-None-Match**. 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 to respond with status code 412 - precondition failed.

Beside other use cases, the **If-None-Match**: header with parameter ***** can be used in a similar way to expose conflicts in resource creation. If any matching entity is found, the operation is supposed 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 definition:

Etag:

name: Etag

description: |

The RFC7232 ETag header field in a response provides the current entity-tag for the selected resource. An entity-tag is an opaque identifier for different versions of a resource over time, regardless whether multiple versions are valid at the same time. An entity-tag consists of an opaque quoted string, possibly prefixed by a weakness indicator.

in: header

type: string

required: false

example: W/"xy", "5", "7da7a728-f910-11e6-942a-68f728c1ba70"

IfMatch:

name: If-Match

description: |

The RFC7232 If-Match header field in a request requires the server to only operate on the resource that matches at least one of the provided entity-tags. This allows clients express a precondition that prevent the method from being applied if there have been any changes to the resource.

in: header

type: string

required: false

example: "5", "7da7a728-f910-11e6-942a-68f728c1ba70"

IfNoneMatch:

name: If-None-Match

description: |

The RFC7232 If-None-Match header field in a request requires the server to only operate on the resource if it does not match any of the provided entity-tags. If the provided entity-tag is `*`, it is required that the resource does not exist at all.

in: header

type: string

required: false

example: "7da7a728-f910-11e6-942a-68f728c1ba70", *

Please also see the section [Optimistic Locking in RESTful APIs](#) for a discussion about alternative approaches.

18. Proprietary Headers

This section shares definitions of proprietary headers that should be named consistently because they address overarching service-related concerns. Whether services support these concerns or not

is optional; therefore, the OpenAPI API specification is the right place to make this explicitly visible. Use the parameter definitions of the resource HTTP methods.

Must: Use Only the Specified Proprietary Zalando Headers

As a general rule, proprietary HTTP headers should be avoided. Still they can be useful in cases where context needs to be passed through multiple services in an end-to-end fashion. As such, a valid use-case for a proprietary header is providing context information, which is not a part of the actual API, but is needed by subsequent communication.

From a conceptual point of view, the semantics and intent of an operation should always be expressed by URLs path and query parameters, the method, and the content. Headers are more often used to implement functions close to the protocol considerations, such as flow control, content negotiation, and authentication. Thus, headers are reserved for general context information ([RFC-7231](#)).

X- headers were initially reserved for unstandardized parameters, but the usage of **X-** headers is deprecated ([RFC-6648](#)). This complicates the contract definition between consumer and producer of an API following these guidelines, since there is no aligned way of using those headers. Because of this, the guidelines restrict which **X-** headers can be used and how they are used.

The Internet Engineering Task Force's states in [RFC-6648](#) that company specific header' names should incorporate the organization's name. We aim for backward compatibility, and therefore keep the **X-** prefix.

The following proprietary headers have been specified by this guideline for usage so far. Remember that HTTP header field names are not case-sensitive.

Header field name	Type	Description	Header field value example
X-Flow-ID	String	The flow id of the request, which is written into the logs and passed to called services. Helpful for operational troubleshooting and log analysis. It supports traceability of requests and identifying request flows through system of many services. It should be a string consisting of just printable ASCII characters (i.e. without whitespace). Verify in a received request that it fits to a specific format, has a sensible maximum length and possibly throw out or escape characters/bytes which could crash your log parsing (line breaks, tabs, spaces, NULL). If a legacy subsystem can only work with flow IDs of a specific format, it needs to define this in its API, or make its own ones.	GKY7oDhpSi KY_gAAAABZ _A

Header field name	Type	Description	Header field value example
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 consumer segment being addressed with a specific product assortment that is offered via CFA retailer catalogs to consumers (see platform glossary [internal link])	52b96501-0f8d-43e7-82aa-8a96fab134d7
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	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 passed-through as generic aspect. Current range is smartphone, tablet, desktop, other	tablet
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-App-Domain	Integer	The app domain (i.e. shop channel context) of the request. Note, app-domain is a legacy concept that will be replaced in new platform by combinations of main CFA concerns like retailer, sales channel, country	16

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 429 with Headers for Rate Limits](#).

Must: Propagate Proprietary Headers

All Zalando's proprietary headers are end-to-end headers. [2: HTTP/1.1 standard ([RFC-7230](#)) 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.]

All headers specified above must be propagated to the services down the call chain. The header names and values must remain unchanged.

For example, the values of the custom headers like **X-Device-Type** can affect the results of queries by using device type information to influence recommendation results. Besides, the values of the

custom headers can influence the results of the queries (e.g. the device type information influences the recommendation results).

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.

19. API Operation

Must: Publish OpenAPI Specification

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 (see [internal narrative](#) for more details). 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

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.

20. Events

Zalando's architecture centers around decoupled microservices and in that context we favour asynchronous event driven approaches. The guidelines in this section focus on how to design and publish events intended to be shared for others to consume.

Events, Event Types and Categories.

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 of the event type registry, and as a validating publish/subscribe broker for event producers and consumers.

Must: Treat Events as part of the service interface

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

Services publishing event data for use by others must make the event schema as well as the event type definition available for review.

Must: Ensure Event schema conforms to Open API Schema Object

To align the event schema specifications to API specifications, we use the Schema Object as defined by the Open API Specifications to define event schemas. This is particularly useful for events that represent data changes about resources also used in other APIs.

The [Open API Schema Object](#) is an **extended subset** of [JSON Schema Draft 4](#). For convenience, we highlight some important differences below. Please refer to the [Open API Schema Object specification](#) for details.

As the Open API 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 definitions](#) 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.

Must: Ensure that Events are registered as Event Types

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.
- 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.

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.

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** can lead to a PATCH or MINOR version revision. MAJOR breaking changes are not allowed.
- Changing an event type with compatibility mode **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 this 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.

The core Event Type structure is shown below as an Open API 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. Note: the name can encode the
        owner/responsible for this EventType and ideally should follow a
        naming pattern that makes it easy to read and understand.
      type: string
      pattern: '[a-zA-Z][-0-9a-zA-Z_]*(\.[a-zA-Z][-0-9a-zA-Z_]*)*'
      example: order.order_cancelled, business_partner.contract
    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
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.

Must: Ensure Events conform to a well-known Event Category

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 used for describing changes to data entities used for data replication based data integration.

The set of categories is expected to evolve in the future.

A category describes a predefined structure that event publishers must conform to along with standard information about that kind of event (such as the operation for a data change event).

The General Event Category.

The structure of the *General Event Category* is shown below as an Open API 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. Previously
    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).

In the example fragment below, the reserved `metadata` field is shown with fields "a" and "b" being defined as part of the custom schema:

Note:

- The General Event in a previous version of the guidelines was called a *Business Event*. 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 the General Event Category to signal steps and arrival points 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 Open API 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. It defines a field called `data` for placing the custom payload information, as well as specific information related to data changes in the `data_type`. In the example fragment below, the fields `a` and `b` are part of the custom payload housed inside the `data` field:

See the following guidelines for more guidance on how to use the Data Change Event Category:

- **Should:** Ensure that Data Change Events match API representations
- **Must:** Use Data Change Events to signal mutations
- **Should:** Use the hash partition strategy for Data Change Events

Event Metadata.

The General and Data Change event categories share a common structure for *metadata*. The metadata structure is shown below as an Open API 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: When the event was created according to the producer.
      type: string
      format: date-time
      example: '1996-12-19T16:39:57-08:00'
    received_at:
      description: |
        When the event was seen by an intermediary such as a broker.
      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: 'JAh6xH40QhCJ9PutIV_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: Ensure that Events define useful business resources

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.

Must: Events must not provide sensitive customer personal data

Similar to API permission scopes, there will be Event Type permissions passed via an OAuth token supported in near future. In the meantime, teams are asked to note the following:

- Sensitive data, such as (e-mail addresses, phone numbers, etc) are subject to strict access and data protection controls.
- Event type owners **must not** publish sensitive information unless it's mandatory or necessary to do so. For example, events sometimes need to provide personal data, such as delivery addresses in shipment orders (as do other APIs), and this is fine.

Must: Use the General Event Category to signal steps and arrival points in business processes

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_ids` 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.

Must: Use Data Change Events to signal mutations

When publishing events that represents created, updated, or deleted data, change event types must be based on the Data Change Event category.

- Change events must identify the changed entity to allow aggregation of all related events for the entity.
- Change events **Should: Provide a means for explicit event ordering**.
- Change events must be published reliably by the service.

Should: Provide a means for explicit event ordering

Some common error cases may require event consumers to reconstruct event streams or replay events from a position within the stream. Events *should* therefore contain a way to restore their partial order of occurrence.

This can be done - among other ways - by adding - a strictly monotonically increasing entity version (e.g. as created by a database) to allow for partial ordering of all events for an entity - a strictly monotonically increasing message counter

System timestamps are not necessarily a good choice, since exact synchronization of clocks in distributed systems is difficult, two events may occur in the same microsecond and system clocks may jump backward or forward to compensate drifts or leap-seconds. If you use system timestamps to indicate event ordering, you must carefully ensure that your designated event order is not messed up by these effects.

Note that basing events on data structures that can be converged upon in a distributed setting (such as [CRDTs](#), [logical clocks](#) and [vector clocks](#)) is outside the scope of this guidance.

Should: Use the hash partition strategy for Data Change Events

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 than 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".

Should: Ensure that Data Change Events match API representations

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: Permissions on events must correspond to API permissions

If a resource can be read synchronously via a REST API and read asynchronously via an event, the same read-permission must apply: We want to protect access to data, not the way data is accessed.

Must: Indicate ownership of Event Types

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: Define Event Payloads in accordance with the overall Guidelines

Events must be consistent with other API data and the API Guidelines in general.

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](#)
- [API Naming](#)
- [Data Formats](#)
- [Common Data Types](#)
- [Hypermedia](#)

Must: Maintain backwards compatibility for Events

Changes to events must be based around making additive and backward compatible changes. This follows the guideline, "Must: Don't Break Backward Compatibility" from the [Compatibility guidelines](#).

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.

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 (note that `x-extensible-enum` is not available in JSON Schema)

Should: Avoid `additionalProperties` in event type definitions

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 "Treat Open API Definitions As Open For Extension By Default" in the [Compatibility](#) section for further guidelines on the use of `additionalProperties`.

Must: Use unique Event identifiers

The `eid` (event identifier) value of an event must be unique.

The `eid` property is part of the standard metadata for an event and gives the event an identifier. Producing clients must generate this value when sending an event and it must be guaranteed to be unique from the perspective of the owning application. In particular events within a given event type's stream must have unique identifiers. This allows consumers to process the `eid` to assert the event is unique and use it as an idempotency check.

Note that uniqueness checking of the `eid` might be not enforced by systems consuming events and it is the responsibility of the producer to ensure event identifiers do in fact distinctly identify events. A straightforward way to create a unique identifier for an event is to generate a UUID value.

Should: Design for idempotent out-of-order processing

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: Follow conventions for Event Type names

Event types can follow these naming conventions (each convention has its own should, must or could conformance level) -

- Event type names must be url-safe. This is because the event type names may appear in URLs published by other systems and APIs.
- Event type names should be lowercase words and numbers, using hyphens, underscores or periods as separators.

Must: Prepare for duplicate Events

Event consumers must be able to process duplicate events.

Most message brokers and data streaming systems offer “at-least-once” delivery. That is, one particular event is delivered to the consumers one or more times. Other circumstances can also cause duplicate events.

For example, these situations occur if the publisher sends an event and doesn’t receive the acknowledgment (e.g. due to a network issue). In this case, the publisher will try to send the same event again. This leads to two identical events in the event bus which have to be processed by the consumers. Similar conditions can appear on consumer side: an event has been processed successfully, but the consumer fails to confirm the processing.

Appendix A: References

This section collects links to documents to which we refer, and base our guidelines on.

OpenAPI Specification

- [OpenAPI Specification](#)

Publications, specifications and standards

- [RFC 3339: Date and Time on the Internet: Timestamps](#)
- [RFC 5988: Web Linking](#)
- [RFC 7159: The JavaScript Object Notation \(JSON\) Data Interchange Format](#)
- [RFC 7230: Hypertext Transfer Protocol \(HTTP/1.1\): Message Syntax and Routing](#)
- [RFC 7231: Hypertext Transfer Protocol \(HTTP/1.1\): Semantics and Content](#)

- [RFC 7807: 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](#)
- [Friboo: utility library to write microservices in Clojure with support for Swagger and OAuth](#)
- [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](#)
- [JSON fields: framework for limiting fields of JSON objects exposed by Rest APIs](#)
- [Tracer: call tracing and log correlation in distributed systems](#)
- [TWINTIP: API definition crawler for the STUPS ecosystem](#)
- [TWINTIP Spring Integration: API discovery endpoint for Spring Web MVC](#)

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.

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 see [the section about usage of the PUT method](#) for a more detailed description on how to update resources via PUT.

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 header 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
> [
>   { code: "00000042"},
>   { code: "00000043"}
> ]

< GET /orders/B00000042

> HTTP/1.1 200 OK
> ETag: osjnfkjbnkq3jlnksjnvkjlsbf
> { code: "B00000042", ... }

< PUT /orders/00000042
< If-Match: osjnfkjbnkq3jlnksjnvkjlsbf
< { code: "00000042", ... }

> 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.

Example:

```
< GET /orders

> HTTP/1.1 200 OK
> [
>   { code: "00000042", etag: "osjnfkjbnkq3jlnksjnvkjlsbf", ... },
>   { code: "00000043", etag: "kjshdfknjqlöwjdsldnfkjbkn", ... }
> ]

< PUT /orders/00000042
< If-Match: osjnfkjbnkq3jlnksjnvkjlsbf
< { code: "00000042", ... }

> 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 a PUT request is made, the version number has to be included in the payload and the server will check that the version in the database is not higher.

Example:

```
< GET /orders

> HTTP/1.1 200 OK
> [
>   { code: "00000042", version: 1, ... },
>   { code: "00000043", version: 42, ... }
> ]
```

```
< PUT /orders/00000042
< { code: "00000042", version: 1, ... }
```

```
> 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

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
> [
>   { code: "00000042", ... },
>   { code: "00000043", ... }
> ]

< PUT /block/00000042
< If-Unmodified-Since: Wed, 22 Jul 2009 19:15:56 GMT
< { code: "00000042", ... }

> 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 *Last-Modified / If-Unmodified-Since* approach or *ETags in result entities*.

Appendix D: Changelog

This change log only contains major changes made after October 2016.

Non-major changes are editorial-only changes or minor changes of existing guidelines, e.g. adding new error code. Major changes are changes that come with additional obligations, or even change an existing guideline obligation. The latter changes are additionally labeled with "Rule Change" here.

To see a list of all changes, please have a look at the [commit list in Github](#).

Rule Changes

- **2018-01-10:** Moved meta information related aspects into new chapter [Meta Information](#).
- **2018-01-09:** Changed publication requirements for API specifications (**Must:** [Publish OpenAPI Specification](#)).
- **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 [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 [Security](#).
- **2016-10-13:** **Should:** [Prefer standard Media type name application/json](#)
- **2016-10-10:** Introduced the changelog. From now on all rule changes on API guidelines will be recorded here.