

# Table of Contents

Introduction	0
Table of Contents	1
Design Principles	2
General Guidelines	3
Security	4
Compatibility	5
Naming	6
Resources	7
HTTP	8
Performance	9
Pagination	10
Hypermedia	11
Data Formats	12
Common Data Objects	13
Common Headers	14
Proprietary Headers	15
API Discovery	16
References	17
Tooling	18



## 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.

## 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.

Furthermore, teams should take part in [API review process \[internal link\]](#).

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



# Table of Contents

# Design Principles

To compare the interface design approaches of SOAP-based web services to those of REST services, the former tend to be centered around operations that are usually use-case specific and specialized. In contrast, REST is all about (business) entities found in the system and exposed as resource endpoints — leading to interfaces that are more broadly usable (here, clients sometimes have to filter out unnecessary information).

## API Design Principles

1. We prefer REST-based APIs with JSON payloads
2. We prefer systems to be truly RESTful
3. 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.”

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

- Fielding Dissertation: [Architectural Styles and the Design of Network-Based Software Architectures](#)
- 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](#)

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.

## General Guidelines

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

### **Must:** API First: Define APIs using OpenAPI YAML or JSON

API First is one of our architecture principles, as per our Architecture Rules of Play. Define, document and review your APIs before delivering them (also see API review procedure).

### **Should:** Provide External Documentation

In addition to defining the API with OpenAPI, it's good practice to provide documentation that describes the API's scope and purpose; architecture and use-case context (including figures); sequence flows; edge cases and possible error situations, etc. Include a link to this documentation using the "externalDocs" property in your RESTful API OpenAPI definition and post it online on GitHub Enterprise pages, on specific team web servers, or as a Google doc.

### **Must:** Write APIs in U.S. English

# 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 your API specification or take a look at the following example.

```
securityDefinitions:
  oauth2:
    type: oauth2
    flow: implicit
    authorizationUrl: https://auth.zalando.com/oauth2/access_token?realm=services
    scopes:
      uid: Unique identifier of the user accessing the service.
      fulfillment-order-service.read: Allows to read from the fulfillment order service.
```

In order to see how to define and assign scopes, please refer to the next section.

## Must: Define and assign scopes

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

```
<scope> ::= <standard-scope> |          -- should be sufficient for majority of use cases
           <resource-specific-scope>    -- for special security access differentiation use cases
<standard-scope> ::= <application-id>.<access-type>
<resource-specific-scope> ::= <application-id>.<resource-id>.<access-type>

<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` scope in this case. It's available to every OAuth2 account by default.



# 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 service contracts 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. With Postel's Law in mind, here are some rules for providers and consumers that allow us to make compatible changes without versioning:

## **Should:** Prefer Compatible Extensions

Apply the following rules to evolve RESTful APIs in a backward-compatible way:

- Ignore unknown fields in the payload
- Add optional, non mandatory fields
- Never change and only extend a field's meaning
- 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.
- 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 values can be reduced when used as output 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 (e.g. via casuistic with default behavior).
- Support redirection in case an URL has to change ([301 Moved Permanently](#))

## **Must:** Prepare Clients for Compatible API Extensions (the Robustness Principle)

How to do this:

- Ignore new and unknown fields in the payload (see also Fowler's "[TolerantReader](#)" post)
- Be prepared for new enum values declared with x-extensible-enum (see above); provide default behavior for unknown values, if applicable
- Follow the redirect when the server returns an "HTTP 301 Moved Permanently" response code

## **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 microservice application with a new API (variant)
- create a new API version supported in parallel with the old API by the same microservice

Avoiding versioning also means preferring the first two variants.

## Must: Use Media Type Versioning

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 content schema, not to URI or method semantics.

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

```
Content-Type: application/x.zalando.cart+json;version=2
Accept: 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: [OpenAPI currently doesn't support content negotiation](#), though [a comment in this issue](#) mentions a workaround (using a fragment identifier that gets stripped off).

## 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).

## Should: Provide Version Information in OpenAPI Documentation

Only the documentation, not the API itself, needs version information.

Example:

```
"swagger": "2.0",
"info": {
  "title": "Parcel service API",
  "description": "API for <...>",
  "version": "1.0.0",
  <...>
}
```

During a (possibly) long-running API review phase you need different versions of the API description. These versions may include changes that are incompatible with earlier draft versions. So we apply the following version schema MAJOR.MINOR.DRAFT that increments the...

- MAJOR version, when you make incompatible API changes
- MINOR version, when you add functionality in a backwards-compatible manner
- DRAFT version, when you make changes during the review phase that are not related to production releases

We recommend using the DRAFT version only for unreleased API definitions that are still under review; for example:

```
version 1.4.0 -- current version
version 1.4.1 -- first draft and call for review of API extensions compatible with 1.4.0
version 1.4.2 -- second draft and call for review of API extensions that are still compatible
                  with 1.4.0 but possibly incompatible with 1.4.1
version 1.5.0 -- approved version for implementation and release
version 1.5.1 -- first draft for next review and API change cycle;
                  compatible with 1.4.0 and 1.5.0
```

Hint: This versioning scheme differs in the less strict DRAFT aspect from [semantic version information](#) used for released APIs and service applications.

## Naming

### **Must:** Path segments must be lowercase separate words with hyphens

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:** Query parameters must be snake\_case (never camelCase)

Examples:

```
customer_number, order_id, billing_address
```

### **Must:** JSON field names must be snake\_case (never camelCase)

Intentionally consistent with the query parameters. Examples see above (Query Parameters).

Why: It's essential to establish a consistent look and feel. 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.

### **Must:** You Must Hyphenate HTTP Headers

#### **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 \(RFC 7230\)](#).

### **Could:** Use Standardized Headers

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

### **Must:** Always 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.

## Could: First Path segment May be /api

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.

## Could: 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. See [Embedding resources](#) for more details.

## 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:** 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.

### **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

Basic URL structure:

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

Examples:

```
/carts/1681e6b88ec1/items  
/carts/1681e6b88ec1/items/1
```

### **Could:** 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 is a collection for sales orders; however, sales orders have globally unique id and some services may choose to access the orders directly, for instance:

```
/customer/1681e6b88ec1  
/sales-order/5273gh3k525a
```

## Should: Limit of Resources

To keep maintenance manageable, and to avoid violations of the “separation of concern” principle, an API should not expose more than 16 resources. If you exceed 16 resources, first check if you can split them into separate subdomains with distinct APIs.

## Should: Limit 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  $\leq 3$  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)

# HTTP

## **Must:** Use HTTP Methods Correctly

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

### **GET**

- reads a resource or set of resource instances, respectively
- individual resources will usually generate a 404 if the resource does not exist; collection resources may return either 200 or 404 if the listing is empty
- must NOT have request body payload

### **PUT:**

- fully uploads an entity, i.e. provides a complete replacement by the resource representation passed as payload
- resource instance IDs are maintained by the client (either generated by the client or by the server in a previous POST request) and passed as a URL path segment
- PUT operations are usually only accepted by single resources, not collection resources, as PUT on a collection would imply replacing the entire collection
- usually robust against non-existence of the entity by implicit creation before update

### **PATCH:**

- partial upload, i.e. only a specific subset of resource fields are replaced
- partial resource representation passed as payload has either resource content type with optional fields (to be updated) or a custom content type that also may include instructions of how to change the resource
- usually accepted only by single resources, because the semantics for PATCH on a collection resource are very hard to define
- usually not robust against non existence of the entity

### **DELETE:**

- deletes a resource instance
- DELETE operations are usually only accepted by single resources, not collection resources, as DELETE on a collection would imply deleting the entire collection
- should return either status 404 (Not found) or 410 (Gone) if the resource does not exist

### **POST:**

- creates a resource instance
- resource instance id(s) are created and maintained by server and returned with the output payload
- POST methods should be accepted by collection resources only
- more generally, POST should be used for scenarios that cannot be covered by the other methods. For instance, GET with complex (e.g. SQL like structured) query that needs to be passed as request body payload. In such cases, make sure to document the fact that POST is used as a workaround

### **HEAD**

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

### **OPTIONS**



- returns the available operations (methods) on a given endpoint (usually either as a comma separated list of methods or as a structured list of link templates)
- this operation is rarely implemented, though it could be used to self-describe the full functionality of a resource.

## Must: HTTP Methods must Fulfill Safeness and Idempotency Properties

An operation can be...

- idempotent, i.e. operation will produce the same results if executed once or multiple times (note: this does not necessarily mean returning the same 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 Meaningful 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 Content-Location header. (More details found in the <a href="#">Common Headers section</a> .) <i>Always</i> 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

### 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 - returned if, e.g. when two clients try to create the same resource or if there are concurrent, conflicting updates	PUT, DELETE, PATCH
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	PUT, DELETE, PATCH
422	Unprocessable Entity - semantic error (as opposed to a syntax error which would usually trigger a 400)	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	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 senseful)	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 senseful.	All

All error codes you can find in [RFC7231](#) and [Wikipedia](#) or via <https://httpstatuses.com/>.

## Must: Providing Error Documentation

APIs should define the functional, business view and abstract from implementation aspects. Errors become a key element providing context and visibility into how to use an API. The error object should be extended by an application-specific error identifier since the HTTP status code often is not specific enough to articulate the domain error situation. For this reason, we use a standardized error return object definition — see [Use Common Error Return Objects](#).

The OpenAPI specification shall include definitions for error descriptions that will be returned; they are part of the interface definition and provide important information for service clients to handle exceptional situations and support troubleshooting. You should also think about a troubleshooting board — it is part of the associated online API documentation, provides information and handling guidance on application-specific errors and is referenced via links of the API definition. This can reduce service support tasks and contribute to service client and provider performance.

## Performance

### Should: Reducing Bandwidth Needs and Improving 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 [Support Filtering of Resource Fields](#) below)
- paginate lists of data items (see [Pagination](#) below)
- ETag (and If-[None-]Match) headers to avoid refetch of unchanged resources
- pagination for incremental access of larger (result) lists

Each of these items is described in greater detail below.

### Should: 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/x.person+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/x.person+json;fields=(name,partner(name))

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

The approach we recommend for field filtering is a Zalando Github project, [json-fields](#). It defines a formal grammar for the ANTLR parser generator and provides a ready-to use library for Java / Jackson based projects ([Maven link](#)). Teams that use other JSON serializers are encouraged to contribute to the open source project and create their own parser / framework based on this grammar.

Other approaches we have considered are JSONPath or GraphQL. While they have advantages, neither of them can easily be plugged into an existing serialization process, so they require an additional, manual serialization process, whereas the above solution addresses our main filter use cases and can easily be introduced with a minimum of effort.

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.

## Could: Support the ETag Header

If a resource changes, the contents of the `ETag` header must also change. Combined with the `If-Match` and `If-None-Match` headers, the `ETag` header allows caching of entities.

Its possible contents:

- the entity's version number
- hash of the response body
- hash of the entity's last modified field

Also see [Allow Embedding of Complex Sub-Resources](#) below as contribution to performance optimization as to avoid multiple requests for sub-resources.

# Pagination

## Must: Support Pagination

For client side batch processing and iteration experience, access to lists of data items must support pagination. This holds 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 better for more efficient implementation, especially when it comes to high-data volumes and / or storage in NoSQL databases.

Before choosing offset-based pagination, make sure that it is feasible for efficient realization. Carefully consider the following trade-offs:

- Usability/framework support:
  - offset/limit pagination is more familiar than cursor-based, 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
  - Using offset will create duplicates or lead to missed entries if rows are inserted or deleted, respectively, between fetching two pages
  - 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

This also holds for total count and backward iteration support

Further reading:

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

## Could: Use Pagination Headers Where Applicable

- Set `x-total-count` to send back the total count of entities.
- Set the [link headers](#) to provide information to the client about subsequent paging options.

For example:

```
Link: <http://catalog-service.zalando.net/articles?cursor=62726863268328&limit=100>;  
      rel="next"; title="next chunk of articles"
```

or

```
Link: <http://catalog-service.zalando.net/articles?offset=5&limit=100>;  
      rel="prev"; title="previous chunk of articles"  
Link: <http://catalog-service.zalando.net/articles?offset=205&limit=100>;  
      rel="next"; title="next chunk of articles"
```

Possible relations: `next` , `prev` , `last` , `first`

# Hypermedia

## Should: Use HATEOAS

This is a challenging topic. We don't have specific recommendations yet; we expect them to emerge from trial and error, and from the API Guild's ongoing research. For now, we strongly encourage teams to apply this principle and document results.

Notes on aspects to be analyzed:

- Security concerns with respect to passing OAuth2 tokens
- API discovery by client programmers vs. automated hypermedia driven workflows
- Additional client development work vs. reduced efforts in case of API and state handling changes

Additional resources: [Wikipedia](#), [The RESTful CookBook](#)

## Could: Use URIs for Custom Link Relations

Related or even embedded (sub-) resources should be linked via “meta link attributes” within the response payload; use here the following [HAL](#) compliant syntax:

```
{
  ...
  "_links": {
    "https://docs.team.zalan.do/rels/my-entity": [{
      "href": "https://service.team.zalan.do/my-entities/123"
    }]
  }
}
```

Or the [JSON API](#) compliant one where you also embed the entities directly:

```
{
  ...
  "relationships": {
    "my-entities": {
      "data": [
        { "id": 1, "type": "my-entity" }
      ],
      "links": {
        "related": "/my-entities/123"
      }
    }
  }
}
```

## Should: Use Standards for Hypermedia Link Types, HAL or JSON API recommended

To represent hypermedia links in payload results, we should consider using a standard format like [HAL](#), [JSON API](#), [JSON-LD](#) with [Hydra](#), or [Siren](#). (Hint: Read [Kevin Sookocheff's post](#), for instance, to learn more about the hypermedia types.)

We are in an ongoing discussion in the API guild which standard should be recommended for a certain use case. In the meantime we recommend to use [HAL](#) or [JSON API](#).



## 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 [Conventional Query Strings](#) for naming.

## Must: Modify the Content-Type for Embedded Resources

Embedded resources requires to change the media type of the response accordingly:

```
GET /order/123?embed=(items)
Accept: application/x.order+json
```

```
HTTP/1.1 200 OK
Content-Type: application/x.order+json;embed=(items)
```

## Data Formats

### **Must:** Use JSON as the Body Payload

JSON-encode the body payload. The JSON payload must follow [RFC-7159](#) by having (if possible) a serialized object or array as the top-level object.

### **Must:** Use Standard Date and Time Formats

- Inside the JSON payload, use the date and time formats defined by [RFC 3339](#) — e.g. `2015-05-28T14:09:17+02:00` for a point in time (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](#).
- In HTTP headers (including the proprietary headers), use the [HTTP date format defined in RFC 7231](#).

A zone offset may be used (both, in request and responses) -- this is simply defined by the standards. However, we encourage to restrict to UTC (and go without offsets) when dates are passed and returned. From experience we know that zone offsets are not easy to understand and often not correctly handled. Note also, that zone offsets are different from local times that may include daylight saving time.

When it comes to storage, all dates should be consistently stored in UTC without zone offset. Localization should be done locally by the services that provide user interfaces, if required.

### **Could:** 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](#)

### **Could:** Use Application-Specific Content Types

For instance, `application/x.zalando.article+json`. For complex types, it's better to have a specific content type. For simple use cases this isn't necessary. We can attach version info to media type names and support content negotiation to get different representations, e.g. `application/x.zalando.article+json;version=2`.

# Common Data Objects

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
      format: decimal
      example: 99.95
    currency:
      type: string
      format: iso-4217
      example: EUR
  required:
    - amount
    - currency
```

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

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

## Should: Use Common Address Fields

Address structures play a role in different functional and use-case contexts, including country variances. The address structure below should be sufficient for most of our business-related use cases. Use it in your APIs — and compatible extend it if necessary for your API concerns:

```
address:
  description:
    a common address structure adequate for many use cases
  type: object
  properties:
    salutation:
      type: string
      description: |
        A salutation and/or title which may be used for personal contacts. Hint: not to be confused with the gender in
      example: Mr
    first_name:
      type: string
      description: given name(s) or first name(s) of a person; may also include the middle names
      example: Hans Dieter
    last_name:
      type: string
      description: family name(s) or surname(s) of a person
      example: Mustermann
    business_name:
      type: string
      description: company name of the business organization
      example: Consulting Services GmbH
    street:
      type: string
      description: full street address including house number and street name
      example: Schönhauser Allee 103
    additional:
      type: string
      description: further details like suite, apartment number, etc.
      example: 2. Hinterhof rechts
    city:
      type: string
      description: name of the city
      example: Berlin
    zip:
      type: string
      description: Zip code or postal code
      example: 14265
    country_code:
      type: string
      format: iso-3166-1-alpha-2
      example: DE
  required:
    - first_name
    - last_name
    - street
    - city
    - zip
    - country_code
```

## Must: Use Common Error Return Objects

[RFC 7807](#) defines the media type `application/problem+json` for a common error return object (and the format of the contained JSON object). 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.

Here is the definition in Open API 2.0 YAML format in the case you don't need any extensions:

```
Problem:
type: object
properties:
  type:
    type: string
    format: uri
    description: |
      An absolute URI that identifies the problem type. When dereferenced,
      it SHOULD provide human-readable documentation for the problem type
      (e.g., using HTML).
    example: http://httpstatus.es/503
  title:
    type: string
    description: |
      A short, summary of the problem type. Written in english and readable
      for engineers (usually not suited for non technical stakeholders and
      not localized); example: Service Unavailable
  status:
    type: integer
    format: int32
    description: |
      The HTTP status code generated by the origin server for this occurrence
      of the problem.
    example: 503
  detail:
    type: string
    description: |
      A human readable explanation specific to this occurrence of the
      problem.
    example: Connection to database timed out
  instance:
    type: string
    format: uri
    description: |
      An absolute URI that identifies the specific occurrence of the problem.
      It may or may not yield further information if dereferenced.
required:
- type
- title
- status
```

## **Must:** An error message must not contain the stack trace.

The reasons:

- The API would leak implementation details to the outside world.
- Clients may not rely on pure API implementation details.

# Common Headers

This section describes a handful of headers, which we found raised the most questions in our daily usage.

## Content-Location

This header is used in the response of either a successful read (GET, HEAD) or successful write operation (PUT, POST or PATCH).

In the case of the GET requests it points to a location where an alternate representation of the entity in the response body can be found. In this case one has to set the Content-Type header 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
```

In the case of mutating HTTP methods, the Content-Location header can be used when there is a response body, and then it indicates that the included response body can be found at the location indicated in the header.

If the header value is the same as the location of the created resource (as indicated by the Location header after POST) or the modified resource (as indicated by the request URI after PUT / PATCH), then the returned body is indeed the current representation of the entity making a subsequent GET operation from the client side not necessary.

If your API returns the new representation after a PUT, PATCH, or POST you should include the Content-Location header to make it explicit, that the returned resource is an up-to-date version.

More details in [rfc7231](#)

## 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.

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.	GKY7oDhpSiKY_gAAAABZ_A
X-Uid	String	Generic user id of OpenId account that owns the passed (OAuth2) access token. E.g. additionally provided by OpenIG proxy after access token validation -- may save additional token validation round trips.	w435-dker-jdh357
X-Tenant-Id	String	The tenant id for future platform multitenancy support. <i>Should not be used unless new platform multitenancy is truly supported. But should be used by New Platform Prototyping services.</i> Must be validated for external retailer, supplier, etc. tenant users via OAuth2; details in clarification. Currently only used by New Platform Prototyping services.	Zalando-Fashion-Store
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 <a href="#">platform glossary [internal link]</a> )	101
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

Remember that HTTP header field names are not case-sensitive.

## API Discovery

### **Must:** Applications Must Provide Online Access to Their API (Swagger) Definitions

In our dynamic and complex service infrastructure, it is important to provide a central place with online access to the API definitions of all running applications. All service applications must provide the following two API endpoints:

- endpoint(s) for GET access on its API OpenAPI definition(s), for instance `https://myapi.com/swagger.json`. Either .json or .yaml formats are supported.
- “Twintip” discovery endpoint `https://myapi.com/.well-known/schema-discovery` that delivers the OpenAPI definition endpoint(s) above

Note, these discovery endpoints have to be supported but need not be part of the OpenAPI definition as they do not provide API specific information.

Background: [Twintip](#) is an API definition crawler of the STUPS infrastructure; it checks all running applications via the endpoint above and stores the discovered API definitions. Twintip itself provides a RESTful API as well as an API Viewer (OpenAPI GUI) for central access to all discovered API definitions.

For the time being, this document is an appropriate place to mention this rule, even though it is not a RESTful API definition rule or related to our STUPS infrastructure for application service management.



## 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](#)

# Tooling

## 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
- [Play Swagger](#): build RESTful Play services from OpenAPI specification
- [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

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