smartAPI: Towards a More Intelligent Network of Web APIs

Amrapali Zaveri

Stanford University

Maastricht University

@AmrapaliZ

Web APIs

Application Programming Interfaces (APIs)

ProgrammableWeb

OPEN API INITIATIVE

*Bio*Catalogue

SADI

Find. Integrate. Analyze.

Challenges in Web APIs

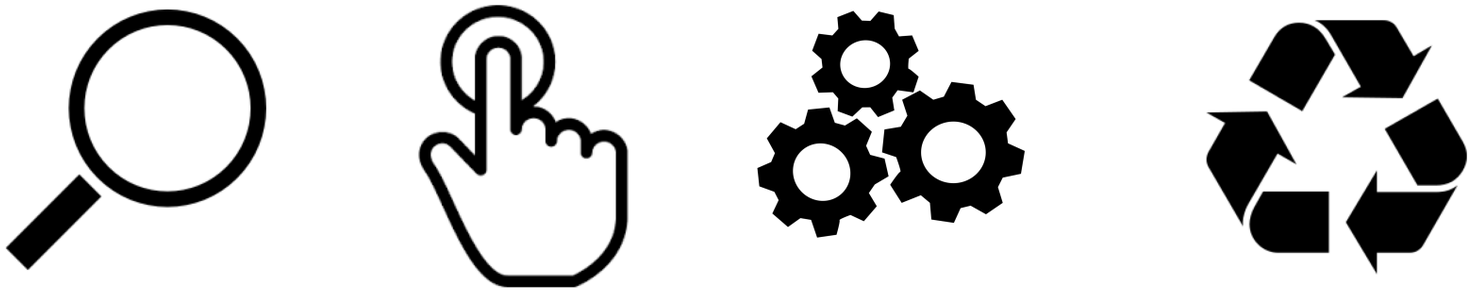
- Lack of rich metadata — creation and reuse
- Lack of tools to semantically annotate APIs
- Unable to automatically discover how APIs connect with each other

Related Work

- Finding APIs
  - collection of useful code, data & documentation
- Reusing APIs
  - server, client-side code, interface, integration
- API annotations for developers
  - WSDL, Swagger (OpenAPI)
- Descriptions for automated clients
  - SAWSDL , OWL-S, (Micro)WSMO, hRESTS, SA-REST, Hydra, RESTdesc, ReLL

smartAPI Ecosystem

Findable Accessible Interoperable Reusable



API providers (authors)

Intelligent (data-driven) annotation of APIs

API Users

A web application to discover APIs and how they connect to each other

The FAIR Guiding Principles for scientific data management and stewardship: https://www.nature.com/articles/sdata201618

API Metadata Survey — in the Wild

Repositories

ProgrammableWeb

17,406 APIs

BioCatalogue

1,187 APIs

elixir Tools and Data Services Registry

6,206 APIs

SADI
Find. Integrate. Analyze.

15,128 APIs

Specifications

OPEN API INITIATIVE

Minimal Information about a Software

schema.org/APIReference

smartAPI specification prototype

Survey of API Metadata — Analysis

- Common metadata elements
- Unique metadata elements
- Coverage
- Cardinality
- Datatype
- Strength of recommendation (MUST, SHOULD, MAY, NOT)

API Metadata Survey — Results

54 Metadata Elements

- MUST - 13
- SHOULD - 19
- MAY – 22
- NOT - 9

Survey available at:
https://goo.gl/wKR51W

5 categories

- API metadata – 20
- Service Provider metadata – 6
- API Operation metadata – 10
- Operation Parameter metadata – 12
- Operation Response metadata – 6

API Metadata Survey — Elements

API Metadata

- Name (Must)
- Access Point (Must)
- Description (Should)
- **Category** (May)

API Operation

- Operation Title (Must)
- **http Method** (Should)
- **Transfer Protocol** (Should)
- **Keywords** (May)

Service Provider

- Responsible Organization (Must)
- Responsible Developer (Must)
- **Funding** (May)
- **Developer Forum** (May)

Operation Parameter

- Operation Name (Must)
- **Parameter Type** (Should)
- **Parameter Value Type** (Should)
- Default Value (May)

Operation Response

- Response Format (Must)
- Example Response Value (Should)
- Response Description (Should)
- **Response Data Type** (May)

smartAPI Specification

W3C Interest Group Note

TABLE OF CONTENTS

1. **Purpose of the smartAPI Specification**
2. **Survey of API Metadata in the Wild**
3. **Audience**
4. **API Metadata Elements**

4.1 Name
4.2 Access Point
4.3 Description
4.4 Response Mime-Type
4.5 External Docs
4.6 Version
4.7 Terms of Service
4.8 SSL Support*
4.9 Authentication Mode
4.10 Category*
4.11 Website
4.12 Publications*
4.13 API Access Restrictions*
4.14 Social Media Links*
4.15 Access Point Mirrors*
4.16 API Metadata Format*
4.17 API Access Mode*
4.18 API Location*
4.19 API Implementation Language*
4.20 API Maturity*

smartAPI Specification

W3C Interest Group Note 13 April 2017

This version:
https://www.w3.org/TR/2017/NOTE-smartAPI-20170413/

Latest published version:
https://www.w3.org/TR/smartAPI/

Editors:
Michel Dumontier, Stanford University, USA
Chunlei Wu, Scripps Research Institute, USA
Amrapali Zaveri, Stanford University, USA
Kathleen Jagodnik, Icahn School of Medicine at Mount Sinai, USA
Ruben Verborgh, Ghent University, Belgium
Raymond Terryn, University of Miami, USA
Paul Avillach, Harvard Medical School, USA
Gabor Korodi, Harvard Medical School, USA

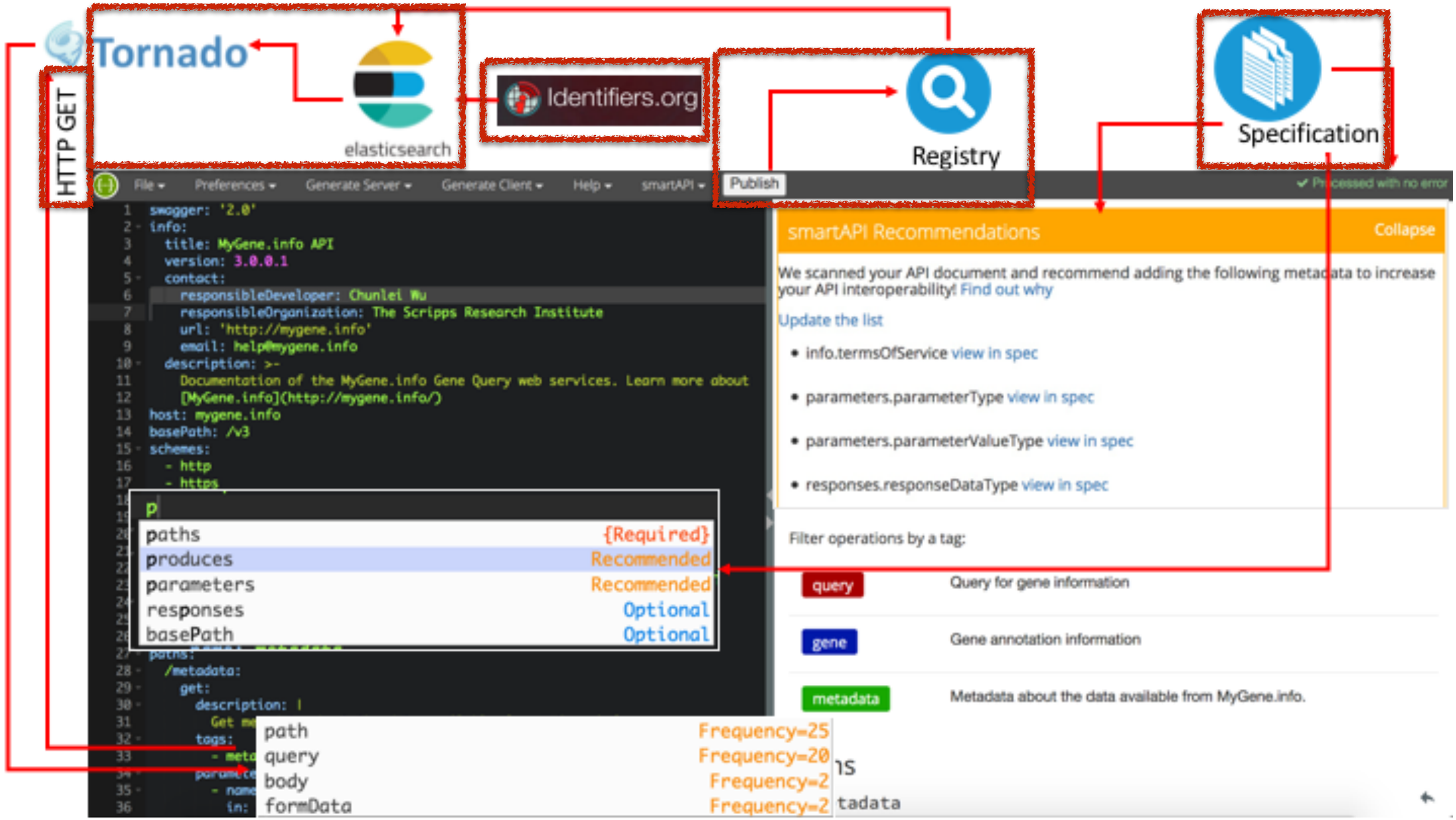
Copyright © 2017 W3C® (MIT, ERCIM, Keio, Beihang). W3C liability, trademark and document use rules apply.

Abstract

This document presents a set of 54 metadata elements (organized into five categories) to usefully describe Web-based Application Programming Interfaces (APIs). These elements were developed by the Big Data to Knowledge (BD2K) API Interoperability Working Group, which conducted a survey of API metadata used in the real world. This group developed the smartAPI Specification as an extension of existing repositories such as Programmable Web, Biocatalogue, and available standards including Open API, schema.org, etc. The aim of the BD2K API Interoperability Working Group is to develop a strategy for maximizing interoperability and reuse of

https://websmartapi.github.io/smartapi_specification/

smartAPI Ecosystem



smart-api.info

Swagger Editor with Open API Validation

File ▾ Preferences ▾ Generate Server ▾ Generate Client ▾ Help ▾ ✔ Processed with no error

```yaml
swagger: '2.0'
info:
  description: >-
    Documentation of the MyGene.info Gene Query web services. Learn more about
    [MyGene.info](http://mygene.info/)
  title: MyGene.info API
  version: 3.0.0.1
host: mygene.info
basePath: /v3
schemes:
  - http
  - https
produces:
  - application/json
tags:
  - description: Query for gene information
    name: query
  - description: Gene annotation information
    name: gene
  - description: Metadata about the data available from MyGene.info.
    name: metadata
paths:
  /metadata:
    get:
      description: |
        Get metadata about the data available from MyGene.info
      tags:
        - metadata
      parameters:
        - name: callback
          in: query
          description: you can pass a "callback" parameter to make a JSONP call
            .
          required: false
          type: string
      operationId: getMetadata
```

MyGene.info API

Documentation of the MyGene.info Gene Query web services. Learn more about MyGene.info

Version 3.0.0.1

Filter operations by a tag:

query	Query for gene information
gene	Gene annotation information
metadata	Metadata about the data available from MyGene.info.

Paths

/metadata

GET /metadata — metadata

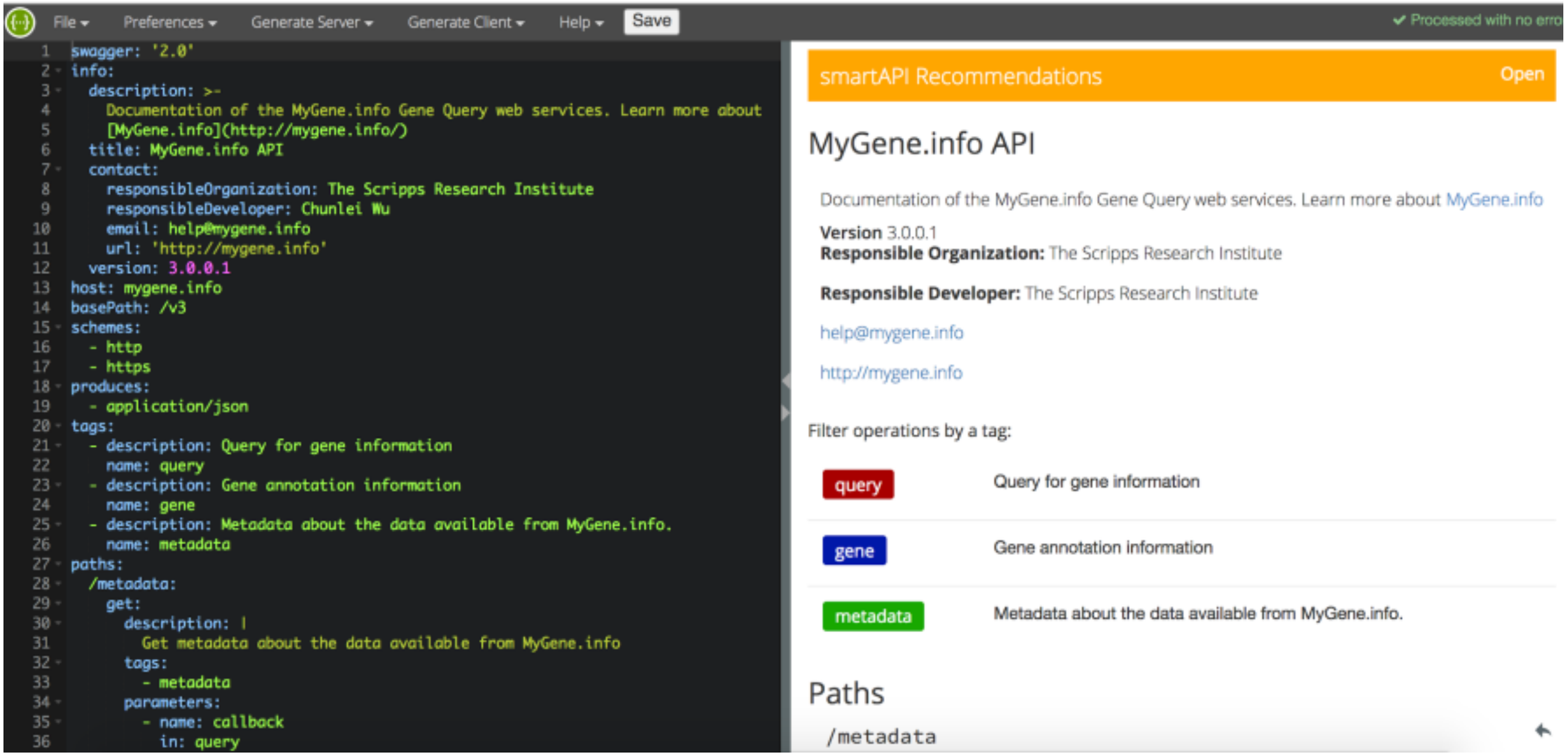
Description

Get metadata about the data available from MyGene.info

Parameters

http://swagger.io/

smartAPI Editor



http://smart-api.info/editor/

smartAPI Editor Features

Field auto-suggestion

```
1  swagger: '2.0'
2
3  info                          {Required}
4  paths                         {Required}
5  definitions                   {Required}
6  produces                      Recommended
7  externalDocs                  Recommended
```

```
1  swagger: '2.0'
2  info:
3    version: "2.0.0"
4    title:  MyGene.info API
5    description: Documentation of the MyGene.info Gene Query web
       services.
6    termsOfService: http://docs.mygene.info/en/v3/terms.html
7    contact:
8      responsibleOrganization: The Scripps Research Institute
9      responsibleDeveloper: The Scripps Research Institute
10     url: http://docs.mygene.info/en/v3/index.html
11     email: help@mygene.info
12
13
```

Value auto-suggestion

```
    operationId: queryGene
    parameters:
      - name: q
        in: query
        description: >-
          Query string. Examples "CDK2", "NM_052827", "204639_at".
            The
          detailed query syntax can be found at
          http://docs.mygene.info/en/latest/doc/query_service.html
        required: true
        type: string
        format: string
      - name: f
```

```
facets                  Frequency=2
fields                  Frequency=2
from                    Frequency=2
fetch_all               Frequency=2
filter                  Frequency=2
dotfield                Frequency=2
species_facet_filter    Frequency=2
```

http://smart-api.info/editor/

Semantic Annotation of Parameter Values

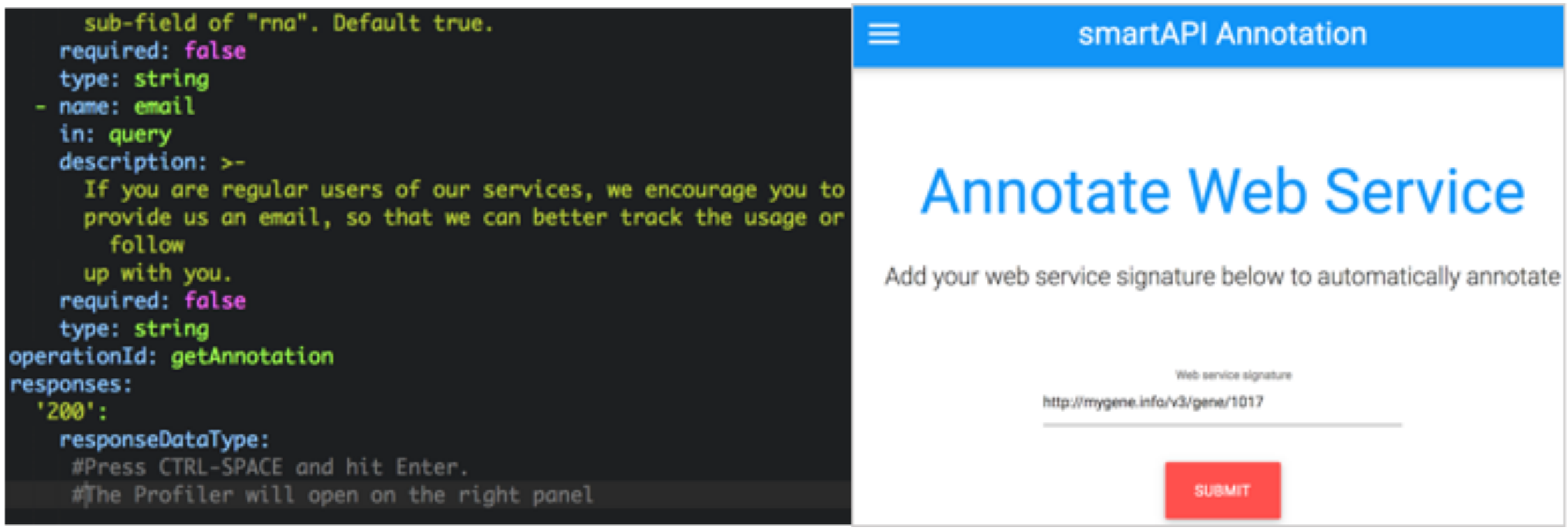
```yaml
parameters:
  - name: geneid
    in: path
    description: >-
      Entrez or Ensembl gene id, e.g., 1017, ENSG00000170248. A re
      Entrez Gene id works too if it is replaced by a new one, e.g
      245794
    required: true
    type: string
    parameterType: InputParameter
    parameterValueType:
      - gene
```

```
http://identifiers.org/ncbigene      Frequency=6
http://identifiers.org/zfin          Frequency=3
http://identifiers.org/hgnc          Frequency=3
http://identifiers.org/xenbase       Frequency=3
http://identifiers.org/hgnc.symbol   Frequency=3
http://identifiers.org/mgi           Frequency=3
http://identifiers.org/unigene       Frequency=3
http://identifiers.org/tair.gene     Frequency=2
http://identifiers.org/gxa.gene      Frequency=0
```

identifiers.org URIs are suggested to semantically annotate parameterValueType

http://smart-api.info/editor/

smartAPI Profiler



The profiler opens within the editor,
once the user wants to annotate response data

http://smart-api.info/editor/

Semantic Annotation of Response Data

```
responseDataType:
  - elementPath: ec
    datatype: http://identifiers.org/ec-code/
  - elementPath: ensembl.gene
    datatype: http://identifiers.org/ensembl/
  - elementPath: ensembl.translation.protein
    datatype: http://identifiers.org/ensembl/
  - elementPath: ensembl.translation.rna
    datatype: http://identifiers.org/ensembl/
  - elementPath: entrezgene
    datatype: http://identifiers.org/ncbigene/
```

responseDataType is automatically annotated by the profiler, within the editor

http://smart-api.info/editor/

smartAPI Registry

http://smart-api.info/registry/

search

Tags

metadata (11) ()

query (11) ()

gene (7) ()

variant (3) ()

PublicDataset (2) ()

SignatureMeta (2) ()

Variant (2) ()

annotation-web-service (2) ()

Schemes

http (25) ()

https (19) ()

Response Formats

application/json (26)

39 results

1. **BD2K PICSURE API v1.0.0**
   As part of the **NIH BD2K PIC-SURE** Center of Excellence http://www.pic-sure.org the HMS Department of Biomedical informatics https://dbmi.hms.harvard.edu is developing an open-source infrastructure that will foster the incorporation of multiple heterogeneous patient level clinical, omics and environment datasets. This system embraces the idea of decentralized datasets of varying types, and the protocols used to access them while still providing a simple communication layer that can handle querying, joining, and computing on.
   details
2. **Broad probabilistic graphical models translator v0.0.1**
   details
3. **Broad probabilistic graphical models translator v0.0.1b**
   details
4. **Broad probabilistic graphical models translator v0.0.1c**
   details
5. **Broad probabilistic graphical models translator v0.0.1a**

smartAPI Registry

http://smart-api.info/registry/

6. **MyGene.info API v3.0.0.1**
   Documentation of the MyGene.info Gene Query web services. Learn more about [MyGene.info](http://mygene.info/)
   details

   version: 3.0.0.1
   organization: The Scripps Research Institute
   developer: Chunlei Wu
   operations:
   - *get /gene/{geneid}*. Retrieve gene annotation based on Entrez or Ensembl gene id, support JSONP and CORS as well.

   input types:
   - http://identifiers.org/ncbigene
   - http://identifiers.org/ensembl

   response types:
   - http://identifiers.org/ec-code/
   - http://identifiers.org/ensembl/
   - http://identifiers.org/ncbigene/
   - http://identifiers.org/pubmed/
   - http://identifiers.org/eco/
   - http://identifiers.org/go/
   - http://identifiers.org/hgnc/
   - http://identifiers.org/homologene/
   - http://identifiers.org/interpro/
   - http://identifiers.org/omim/

smartAPI Use Case

*Given a **gene**, find **other genes** involved in the same **pathway** and find the candidate small **molecules** with these genes as the **perturbation targets***

Input gene

mygene.info

pathway

mygene.info

genes involved in the pathway

mygene.info

smartAPI Use Case

*Given a **gene**, find **other genes** involved in the same **pathway** and find the candidate small **molecules** with these genes as the **perturbation targets***

Input gene → **pathway**

mygene.info → mygene.info

genes involved in the pathway → **list of "active" perturbagens & their associated bioassay dataset IDs**

mygene.info

NIH LINCS PROGRAM

http://lincsportal.ccs.miami.edu/apis/

smart-api.info

smartAPI Use Cases

- MyVariant.info (annotated)
- Rat Genome Database
- BD2K PIC-SURE HTTP API
- Alliance of Genome Resources

img source: http://www.wauchopeshowsociety.com.au/images/sl_website-under-construction.jpeg

Conclusions

- smartAPI **Specification** — 54 API metadata elements to describe APIs

- smartAPI **Editor** — composition, semantic annotation and validation of API metadata

- smartAPI **Profiler** — automatically annotate the API response data with semantic identifiers

- smartAPI **Registry** — submit, search, and browse API descriptions and obtain field-specific metadata suggestions

smart-api.info

Future Work

- Upgrade to openAPI v3
- Export functionality to other repositories
- Providing API descriptions as JSON-LD
- Improving registry interface

smart-api.info

Acknowledgements

NIH Commons Big Data to Knowledge (BD2K) API Interoperability Working Group

- Michel Dumontier
- Shima Dastgheib
- Chunlei Wu
- Trish Whetzel
- Ruben Verborgh
- Paul Avillach
- Gabor Korodi
- Raymond Terryn
- Kathleen Jagodnik
- Pedro Assis
- Gregg Kellogg
- Mark Wilkinson
- Jeff De Pons
- Denise Luna

smart-api.info

Thank You!
Questions?

smart-api.info

Maastricht University

amrapali.zaveri@maastrichtuniversity.nl

@AmrapaliZ