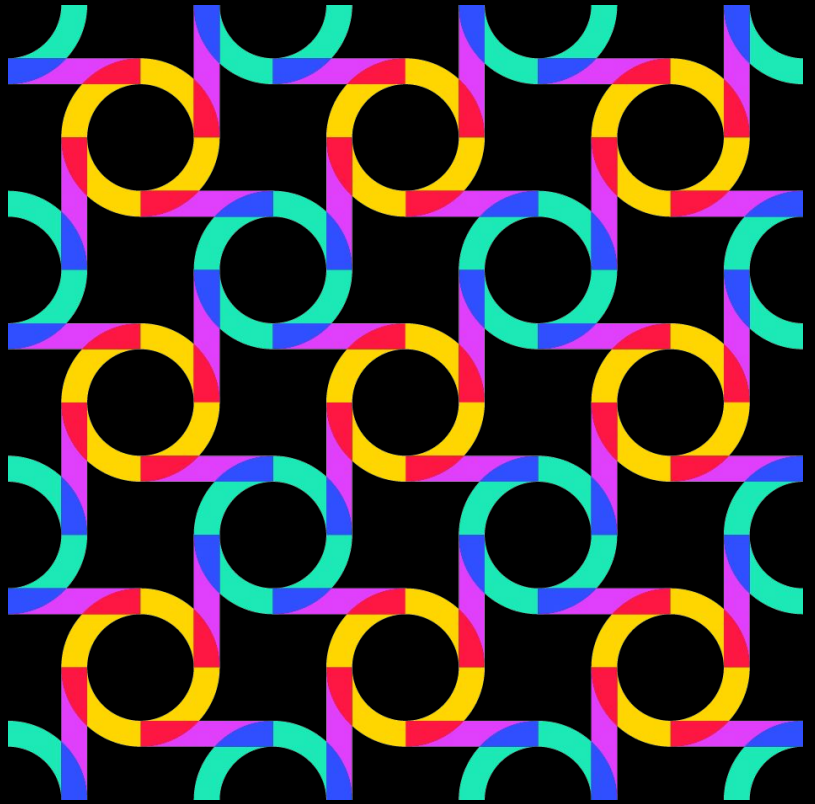
Introduction to Smart Data Models



FIWARE Foundation

Data Modelling Expert

Alberto Abella

i4Trust

i4Trust Website

i4Trust Community

FIWARE FOUNDATION

iSHARE

FundingBox

Contents and goals

- About:
  - During this part of the session you will get introduced to the Smart Data Models initiative
- This session will:
  - Explain basics about the Smart Data Models initiative and how it is governed
  - Introduce you to data models for specific verticals that are already available
  - Explain how you may contribute extensions to existing data models under Smart Data Models
  - Explain how you may contribute new data models under Smart Data Models
- Goals:
  - After this session you will be able to implement a service using a Smart Data Model
  - You will be able to explain how to become an active contributor to the Smart Data Model initiative
- Target Audience:
  - LEBDs
- Link to the content

Smart Data Models: Introduction and Governance

i4Trust

Introduction and Governance

Steering Board :

- Three members (April-2021)

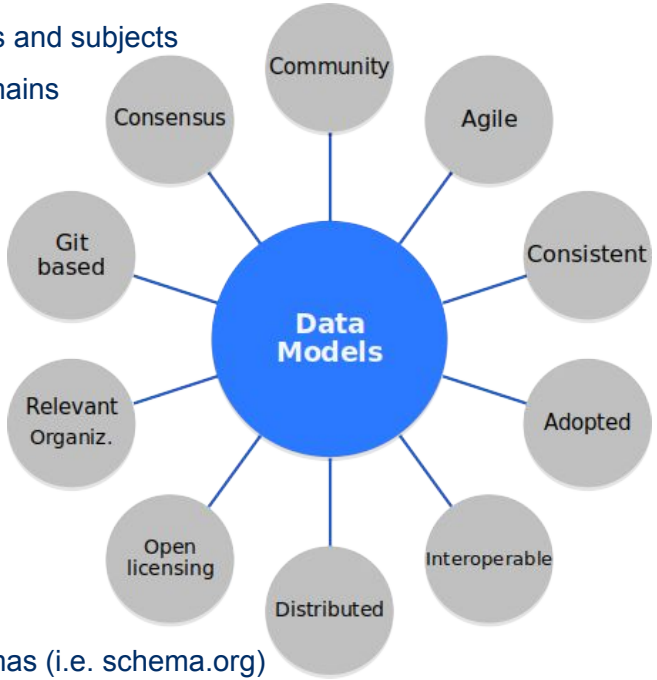
- IUDX: Indian Urban Data Exchange
  - Public entity supporting data interchange for Smart cities in India

- TMForum:
  - Worldwide Telco association

- FIWARE Foundation
  - Curator of the open source FIWARE platform and its ecosystem

Introduction and Governance

- A **community** site with **detailed data models** available for open use for multiple sectors
- Together with other **relevant organizations** in the curation of the different domains and subjects
- Providing **coherence and consistency** between data models across different domains
- To create a method for **AGILE standardization** and evolution these data models
- To provide extended usefulness to FIWARE platform users in terms of:
  - Extended **interoperability**
  - Reduced time dedicated to data model codying
  - Accumulated experience tested in real case scenarios
  - Mapped to be integrated with other platforms
- Using **open licensing** to allow extensive use and adoption
- Used in **real** case scenarios (and based on real use cases)
- Based on git platform and github as development frontend
- Consensus as main decision method
- Based on **widely adopted** standards (including ontologies and international schemas (i.e. schema.org)

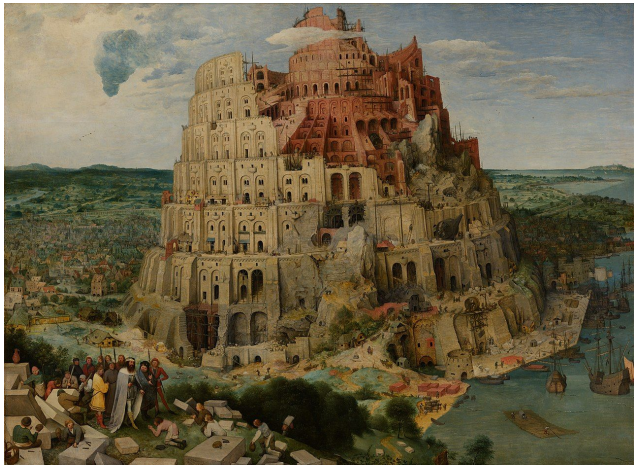


i4Trust

Introduction and Governance

Differential advantages

1. **Agile standardization.** Standardization time takes Weeks vs months/years
2. **Easy contribution.** One single file source of truth.
3. **Based on actual experience**. All data models are based on real case scenario.
4. **Multilanguage**. Automation translation of specs.
5. **1 change for all domains**. Changing a data model impact on all the domains related to it.



By Pieter Brueghel the Elder - bAGKOdJfvfAhYQ at Google Cultural Institute, zoom level maximum, Public Domain, https://commons.wikimedia.org/w/index.php?curid=22178101

Smart Data Models: Structure and contents. Verticals

i4Trust

Smart Data Models: Structure and contents. Verticals

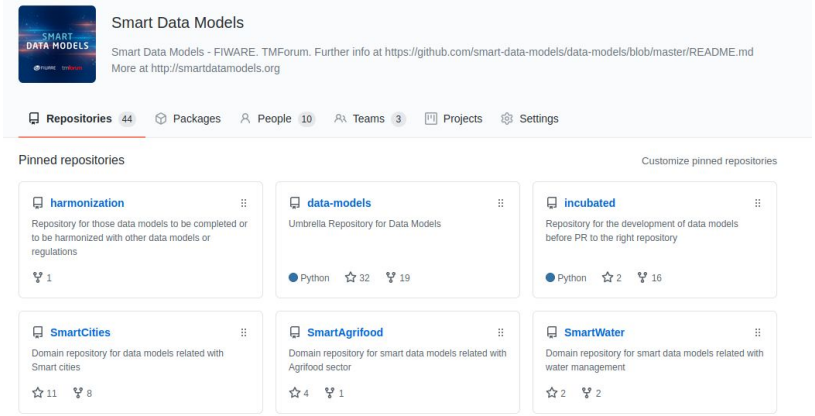
- What is a Smart Data Model:
  - It is the combination of 3 elements
    - A technical description of the data types and relationships of the attributes of an element that has to be modeled
    - A documented description of these attributes aligned with the technical description
    - Some examples of the use of the data model
  - Based on real case scenario (not theory or academic research)
  - With an open license allowing its use, share and modification

- What is NOT a Smart Data Model
  - It is not an ontology describing the elements of an area of knowledge
  - A new standard
  - An academic exercise

Structure and contents. Verticals

GITHUB

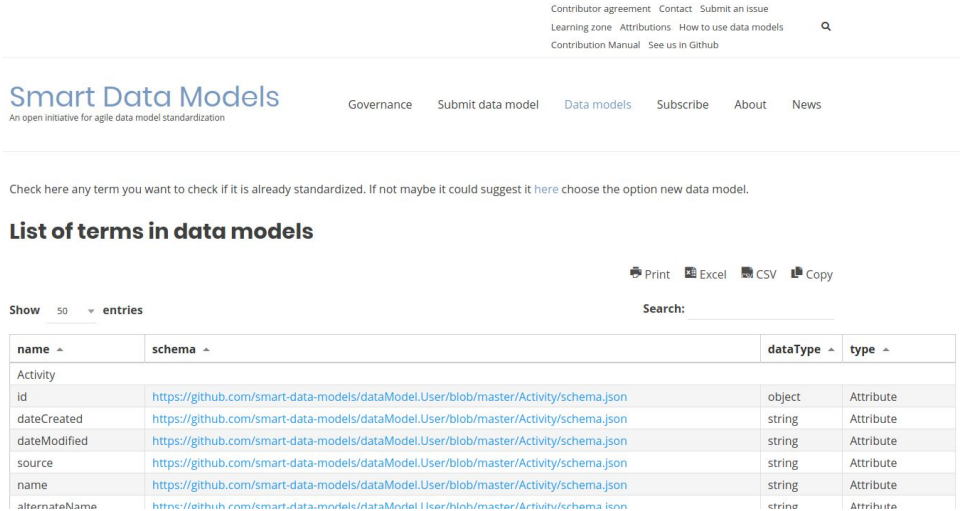
http://github.com/smart-data-models



- Oriented to **developers**
- All resources available
- Contribution by PR
- Issues on data models

SITE (wp)

http://smartdatamodels.org



- Oriented to **end users**
- News on updates (subscription)
- Check attributes and enumerations
- Check descriptions

i4Trust

Structure and contents. Verticals

DATA-MODELS

- Guides for coding new data models
- Template for new data models and examples
- Directory for scripting tools to check data models
- Inventory of domains and data models
- Inventory of attributes and terms

data-models
Umbrella repo

DOMAINS REPOSITORIES

Readme
pointing to the list of subjects
General info or shared resources

Smart Water | Smart Cities | Smart Environment | Smart Building | Smart Destinations | Smart Agrifood | Smart Robotics | Smart Manufacturing | Smart Energy

DATA MODELS

README.md
/doc/spec.md
/examples
schema.json
Current Adopters

Subject 1 (sewage) | Subject 2 (parking) | Subject 3 (weather) | Subject 4 (...)

SUBJECTS' REPOSITORIES

Readme pointing to the list of data models for the objects
Contributors.md
subject-schema.json

LIFECYCLE MANAGEMENT REPOSITORIES

Incubated | Harmonization

i4Trust

Structure and contents. Verticals

About: 554 data models available. Data Models per domain*:

- 'SmartEnergy': 414
- 'SmartCities': 59
- 'CrossSector': 33
- 'SmartWater': 22
- 'SmartAgrifood': 19
- 'SmartEnvironment': 16
- 'SmartDestination': 10
- 'SmartAeronautics': 6
- 'Smart-Sensoring': 4
- 'SmartRobotics': 1

Last Subjects :

- dataModel.Forestry
- dataModel.SocialMedia,
- dataModel.WasteWater,
- dataModel.OPCUA,
- dataModel.OpenChannelManagement,
- dataModel.QueueManagement

*Some subjects are linked to several domains so the # of data models exceeds the actual figure

** ROS data models could mean > 100 data models

Updated 25-4-2021

Structure and contents. Verticals

Incubated repository (They could end up being official data models or not):

Total estimated : 98 (+49)

In domains of:

SmartRobotics 53* (not related to ROS)

SmartCities 12

SmartWater 11

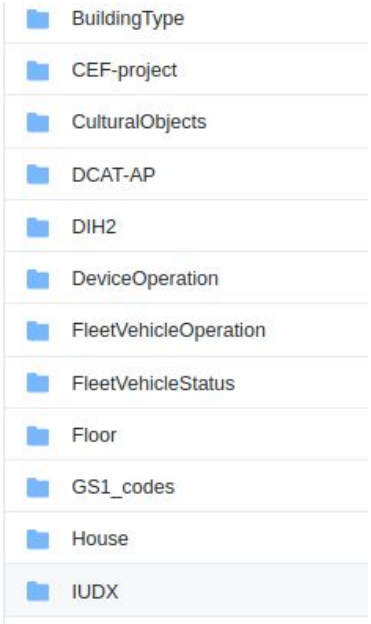
CrossSector 11

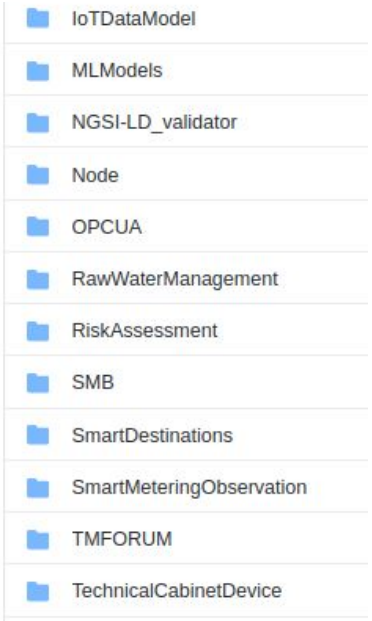
Smart-Sensoring 6

Smart Environment 3

SmartAgrifood 1

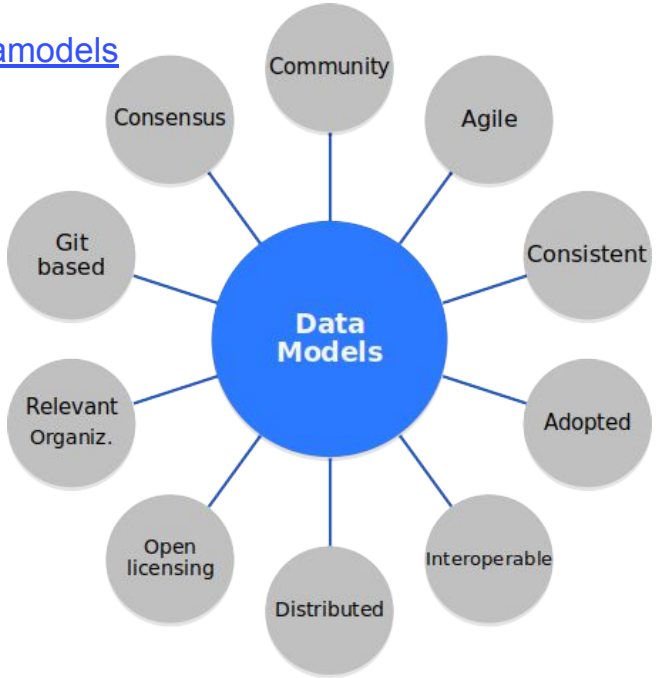
Cross Sector 5





Structure and contents. Verticals

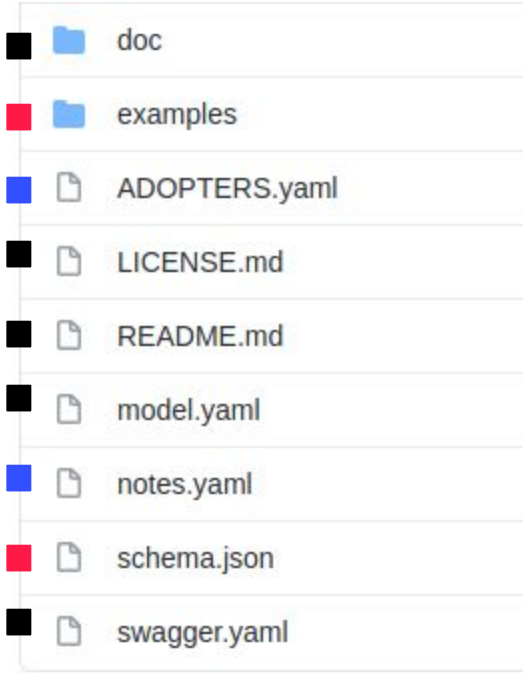
- Services for users
  - News, subscription to newsletter, and twitter account @smartdatamodels
  - Search data models
  - Repositories of data models, quick list.
- Services for developers
  - Checklist for submitting your new data model
  - Create your data model with a spreadsheet
  - Create your data model in the incubated repository
  - Create a examples compliant with a data model (NGSI-LD)
  - Validate if your data model is well documented
  - Validate your payload (external)



i4Trust

Structure and contents. Verticals

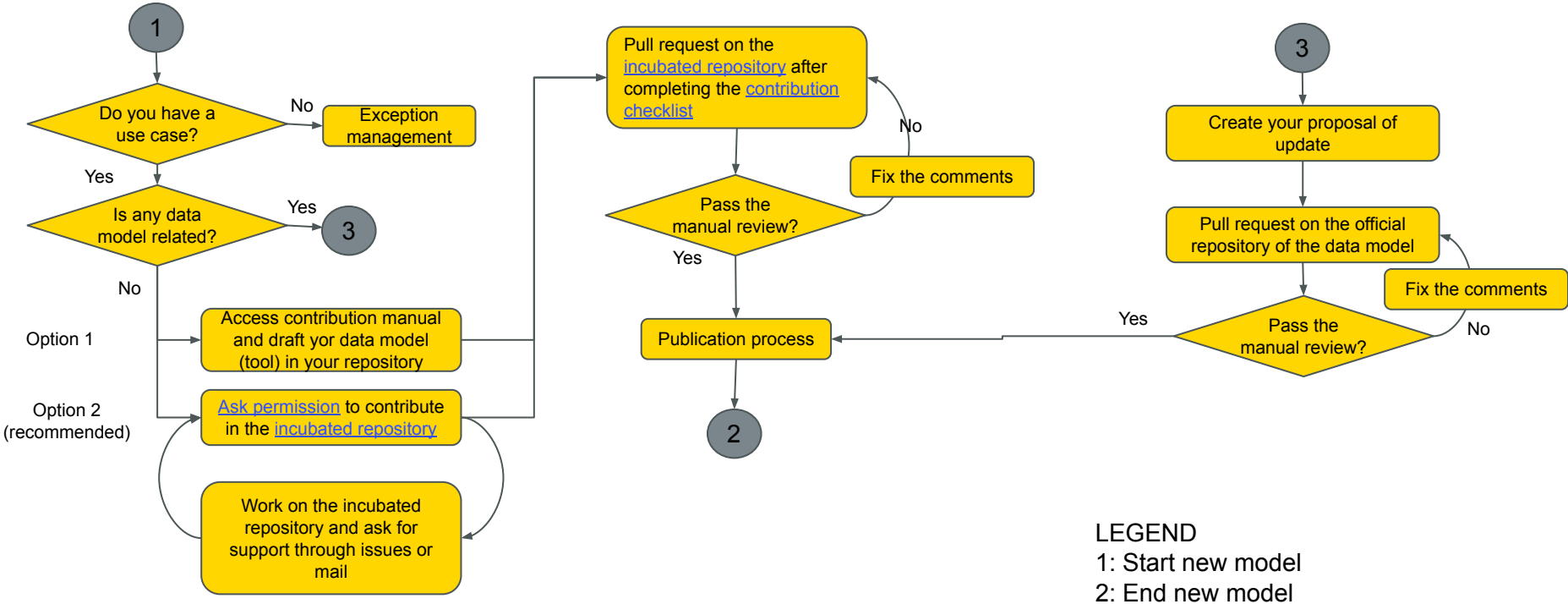
- Contents:
  - *doc*: directory for specifications
  - *examples* : Directory for examples
  - *ADOPTERS.yaml* use cases of the data model
  - *LICENSE.md* license of the data model. I.e. CC-BY
  - *README.md* pointer to the main elements of the data model, including examples, specifications and other services
  - *model.yaml* technical description of the attributes for embedding into the specifications
  - *notes.yaml* additional file for customization contents for the specifications
  - *schema.json* single source of truth of the model. Validates only the key-values payloads
  - *swagger.yaml* yaml file required for the interactive specification and future test of services



Smart Data Models: Become a contributor.

i4Trust

Smart Data Models: Become a contributor.



LEGEND
1: Start new model
2: End new model
3: Start update existing model

i4Trust

Smart Data Models: Exercise
Turn a data source definition into a data model

i4Trust

Exercise

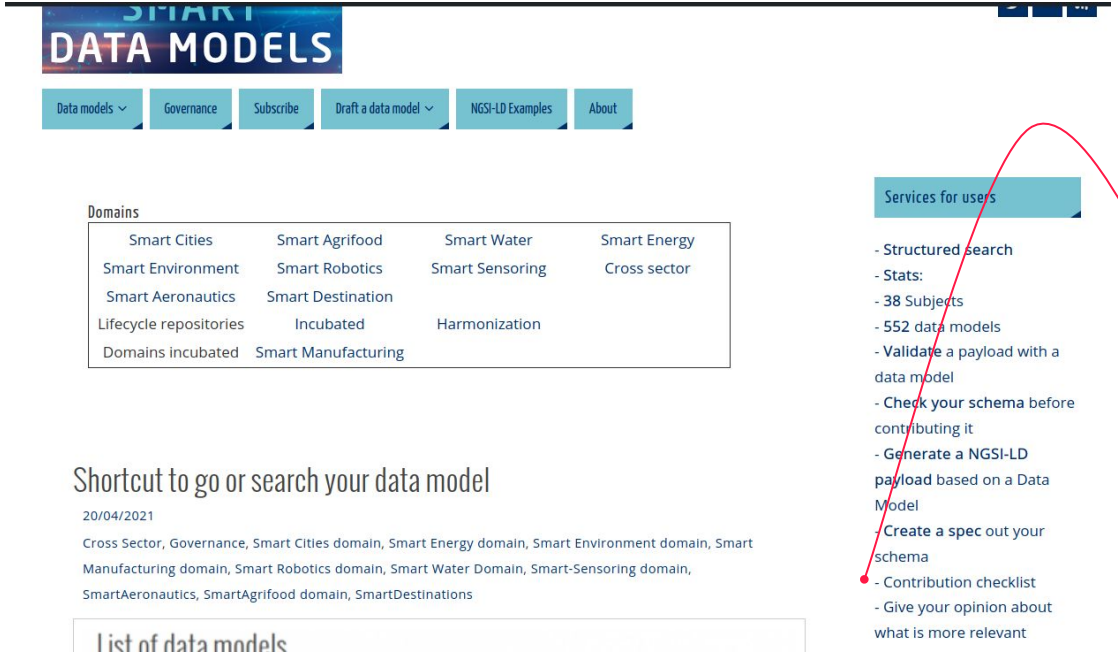
- Expected knowledge from the participants
  - Knowledge of NGSI-v2 and NGSI-LD and their differences
  - Some work with a Context Broker (any of them)
  - JSON and JSON Schema
  - Git and GitHub concepts
  - A code editor like PyCharm

Exercise

- It is required **your github users** to be written here (http://bit.ly/github_users) to grant you access to the incubated repository.
  - If you do not have a github user, then go here: http://bit.ly/register_github
- Use the repository incubated. https://github.com/smart-data-models/incubated/tree/master
- The final exercise is to **submit a complete data model.**
- Complete the creation of an official data model through all its steps.
- It will be done with official sources, so the result of the exercise, if completed, will become an official data model of the initiative.
- Comments to the contribution manual will be incorporated
- If not completed during session time It can be completed afterwards

Become a contributor

- Checklist for contribution



i4Trust

Exercise

Data sources

These data sources have dozens of properties. We'll only take a few for the exercise.

- https://www.schema.org/MedicalCondition
- https://www.schema.org/MedicalGuideline
- https://www.schema.org/Drug
- https://www.schema.org/MedicalScholarlyArticle
- https://www.schema.org/LocalBusiness
- **https://www.schema.org/Organization**
- https://www.schema.org/Restaurant

- Other available. https://www.schema.org/docs/health-lifesci.home.html

Take a minimum of 5 properties, one an object / array.

Exercise

Data sources

Other data sources could be valid as well as:

1. It has documented attributes
2. It has clear data types for the attributes
3. It is on use in some real case scenario

Exercise

- Steps:
  1. Review the contribution manual
  2. Access to the data source
  3. Open our spreadsheet
  4. Paste the data definitions in the spreadsheet according to the contribution manual.
  5. Generate the json schema
  6. Validate the json schema
  7. Generate the example of payload
  8. Validate the example against the schema
  9. Submit the new data model

Exercise

- Review the contribution manual

1. Contribution manual is linked in the main page or in https://bit.ly/contribution_manual

Exercise

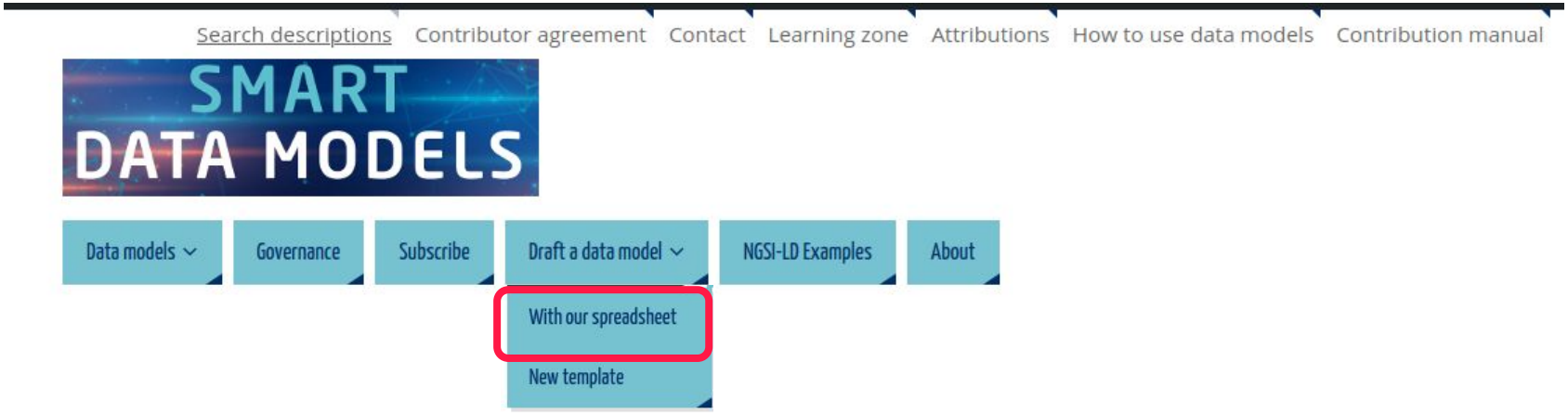
- Access to the data source. Examples available.

a. https://www.schema.org/MedicalCondition
b. https://www.schema.org/MedicalGuideline
c. https://www.schema.org/Drug
d. https://www.schema.org/MedicalScholarlyArticle
e. https://www.schema.org/LocalBusiness
f. https://www.schema.org/Organization
g. https://www.schema.org/Restaurant

h. https://www.schema.org/docs/health-lifesci.home.html

Exercise

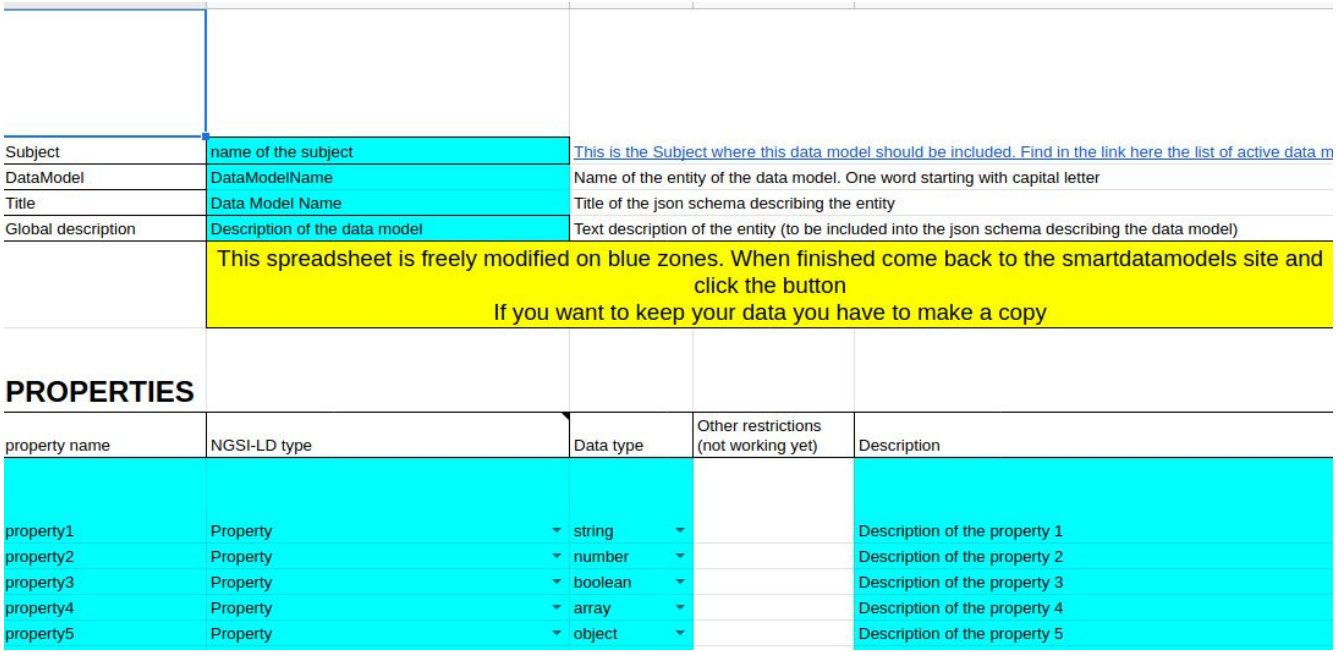
- Open our spreadsheet (need a google drive account)



i4Trust

Exercise

- Paste the data definitions in the spreadsheet according to the contribution manual.



Subject	name of the subject	This is the Subject where this data model should be included. Find in the link here the list of active data m
DataModel	DataModelName	Name of the entity of the data model. One word starting with capital letter
Title	Data Model Name	Title of the json schema describing the entity
Global description	Description of the data model	Text description of the entity (to be included into the json schema describing the data model)

This spreadsheet is freely modified on blue zones. When finished come back to the smartdatamodels site and click the button
If you want to keep your data you have to make a copy

PROPERTIES

property name	NGSI-LD type	Data type	Other restrictions (not working yet)	Description
property1	Property	string		Description of the property 1
property2	Property	number		Description of the property 2
property3	Property	boolean		Description of the property 3
property4	Property	array		Description of the property 4
property5	Property	object		Description of the property 5

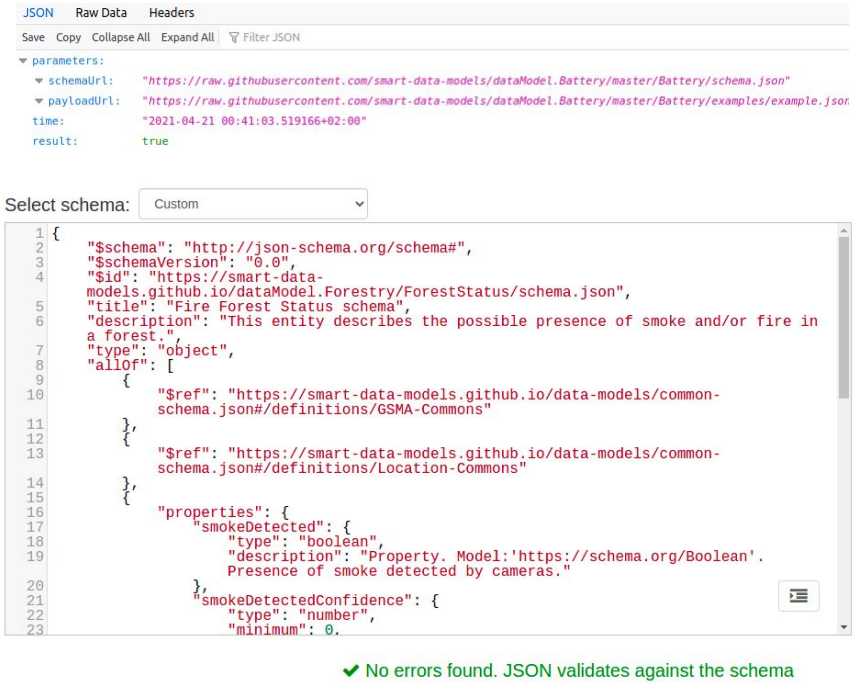
Exercise

- Generate the json schema
  - Fix possible errors
  - Provide more detail / limits
  - Include Units
  - Include Enumerations
  - Include Privacy
  - Include limitations to values
  - Required Properties

i4Trust

Exercise

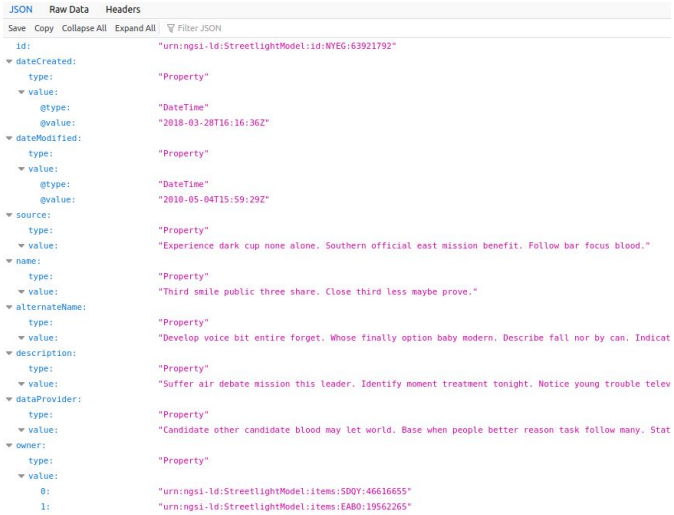
- Validate the json schema
  - Validation of the json schema https://www.jsonschemavalidator.net/
  - Validation of the documentation https://smartdatamodels.org/index.php/data-models-contribution-api/



i4Trust

Exercise

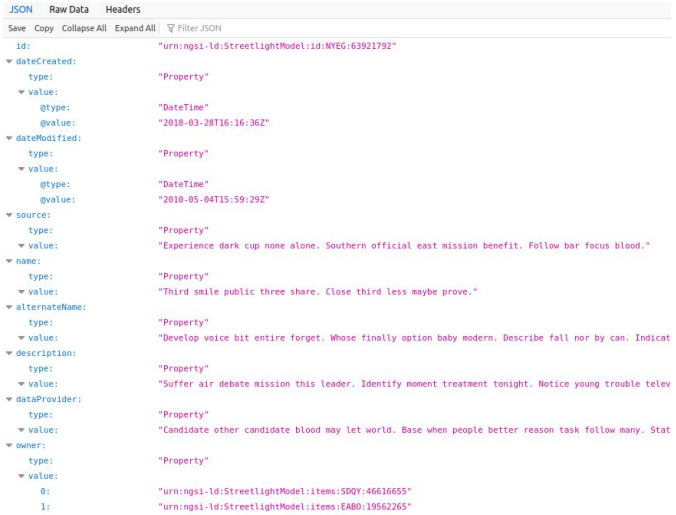
- Generate the normalized example of payload
  - I.e. NGSI-LD is available
- Needs edit for meaningful data



i4Trust

Exercise

- Generate the key-values of payload
  - Use the script or manually with your code editor



i4Trust

Exercise

- Validate the payload (**keyvalues**). https://www.jsonschemavalidator.net/

Select schema: Custom

```
{
    "$schema": "http://json-schema.org/schema#",
    "$schemaVersion": "0.0",
    "$id": "https://smart-data-
    models.github.io/dataModel.Forestry/ForestStatus/schema.json",
    "title": "Fire Forest Status schema",
    "description": "This entity describes the possible presence of smoke and/or fire in
    a forest.",
    "type": "object",
    "allOf": [
        {
            "$ref": "https://smart-data-models.github.io/data-models/common-
            schema.json#/definitions/GSMA-Commons"
        },
        {
            "$ref": "https://smart-data-models.github.io/data-models/common-
            schema.json#/definitions/Location-Commons"
        },
        {
            "properties": {
                "smokeDetected": {
                    "type": "boolean",
                    "description": "Property. Model:'https://schema.org/Boolean'.
                    Presence of smoke detected by cameras."
                },
                "smokeDetectedConfidence": {
                    "type": "number",
                    "minimum": 0,
```

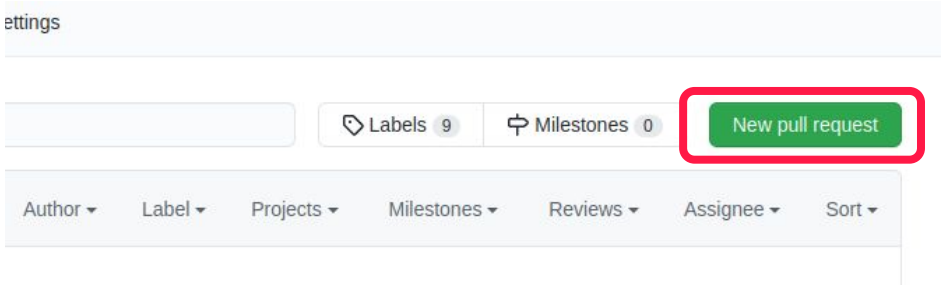
✔ No errors found. JSON validates against the schema

Input JSON:

```
{
    "id": "FireForestStatus-South-1",
    "dateObserved": "2021-02-24T00:00:00Z",
    "location": {
      "type": "Point",
      "coordinates": [42.206302, -7.887465]
    },
    "name": "Ourense Forest - South",
    "description": "Status of the Ourense Forest (south)",
    "refDevice": {
        "type": "Relationship",
        "object": ["ground-humidity-sensor:1"]
    },
    "smokeDetected": false,
    "smokeDetectedConfidence": 0.9,
    "fireDetected": false,
    "fireDetectedConfidence": 0.8,
    "fireRiskIndex": 0.1,
    "litterCoverage": 0.6,
    "relativeHumidity": 0.70,
    "soilTemperature": 25
}
```

Exercise

- Fork the incubated repository
- PR your changes to the data model



They will appear in the front page in 5 minutes maximum (right column, bottom)

PR and issues

updated every 5 minutes (if empty, it means we are updating, plese refresh)
dataModel.WasteManagement
WasteContainer (storedWasteKind) enum, other values?
dataModel.Transportation
remove minimum from amperage and voltage
dataModel.Battery
Possible inconsistency in rechargeTime attr
dataModel.AutonomousMobileRobot
Added robotics data models for autonomous mobile robots
incubated
commit
added smb to OrionBroker
dataModel.SocialMedia
Pull request to fix inconsistencies

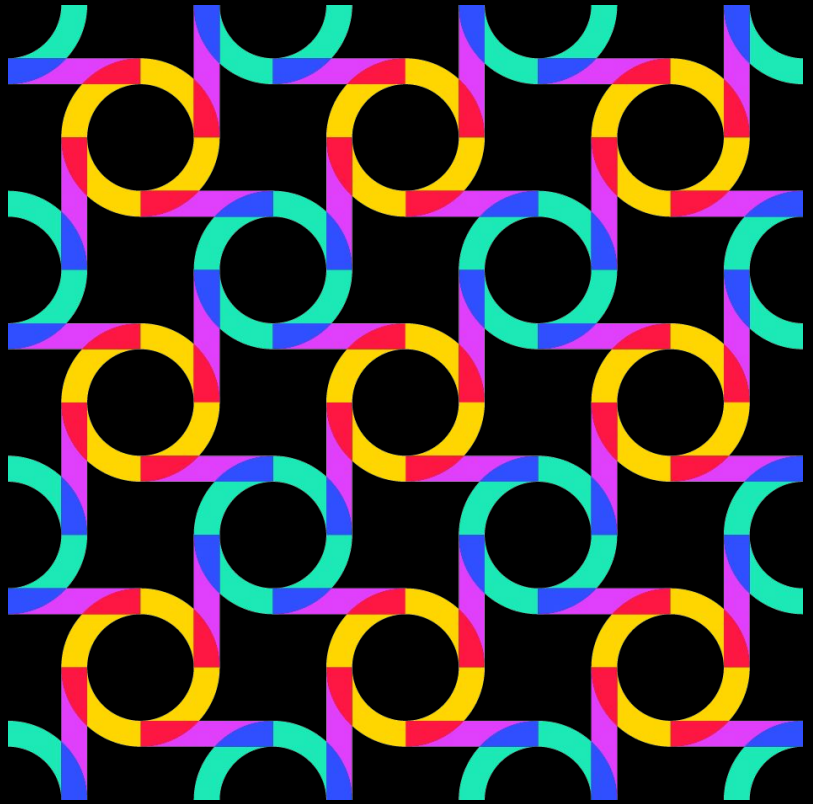
Smart Data Models: Summary

i4Trust

Summary

- Goal
  - Allow real interoperability between NGSI data sources
- Contribution:
  - Always possible as long as it has a use case, and comply with contribution workflow
- Use of the data models
  - Search tools for finding the right data model
  - Best not to reinvent the wheel
- Differential advantages of agile standardization
  - Quick answer
  - Do not invent
  - Easy contribution
  - Single source of truth
  - Better simple and useful than technically ‘correct’ and powerful

Thank you!



i4Trust

i4Trust has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement no 951975.