
Open Data Services Sphinx Base

Release 0.0

Open Data Services

Dec 03, 2020

Contents

1	Contents	1
2	Background	3
3	Status of the data model	5
4	Overview of the data model	7
5	Key features	9
5.1	Modelling declarations	9
5.2	Flexibility	9
5.2.1	Data profile for 'Employment and earnings' interest	9
5.3	Open codelists	10
6	Outstanding issues	11
6.1	Publisher metadata and republishing	11
6.2	Change over time.	11
6.3	Packaging and publication format	12
7	About	13

CHAPTER 1

Contents

In early 2020, [Open Data Services Co-operative](#) and [OpenDemocracy](#) collaborated on a project to prototype a searchable, public database of politicians' and elected members' interests. This resulted in [declared.info](#).

Declarations of interest were obtained by scraping websites of a representative range of UK political bodies. The diversity of formats, detail and quality represented by these declarations made work on [declared](#). unnecessarily challenging, and the site itself not as useful as it might be. We also examined registers of interest provided in response to Freedom of Information requests, but did not import them into [declared](#)..

The data from all these sources varies widely in quality and format, meaning it cannot be easily merged or compared with other datasets. Public declarations of potential conflicts of interest are a required component of the checks and balances in a democratic society. When these declarations only have value in isolation and cannot easily be compared with others, or linked to other types of information, they have limited value. This is especially true at a time when private sector and non-democratic actors face no such limitations.

If democratic institutions and bodies adopted an open data standard for publishing their registers of interests, tools such as [declared.info](#) could be built cheaply and have far-reaching scope. The public interest would be far better served: citizens would be able to hold elected members to account; journalists would be able to analyse and report on the influence of organisations; the operations of politicians and organisations would be more transparent.

Status of the data model

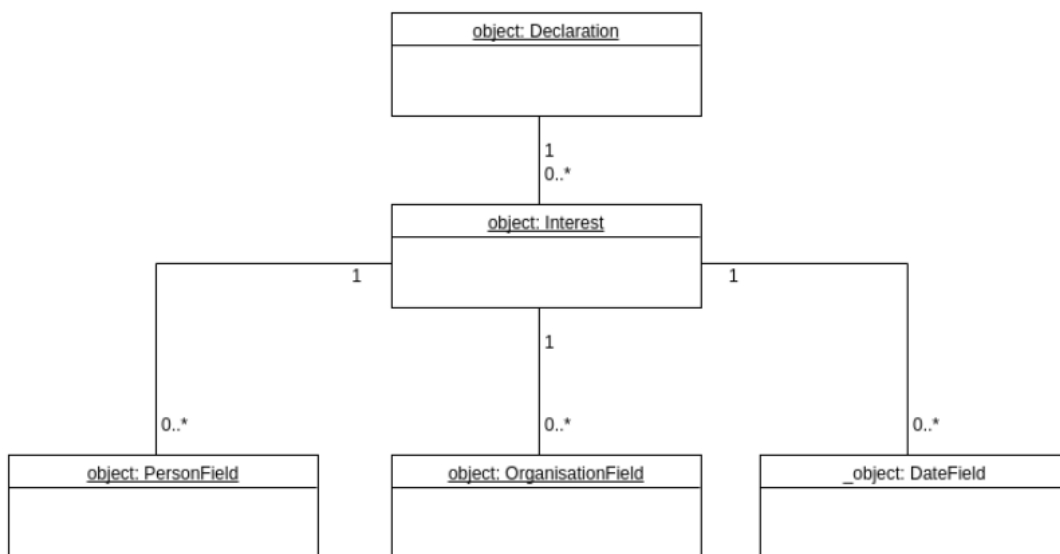
We present here the beginnings of an open data standard: a draft data model to represent information declared by people about their interests. The structure of the data (the data model) is informed by:

- the experience of dealing with unstructured, variously formatted information in this domain;
- an understanding of how tightly the format of declarations can be tied to regulatory requirements and language;
- a recognition that an 'interest' may cover an extremely wide range of activities, roles and engagements;
- an assumption that the added value of a structured format is in enabling patterns across declarations to be more easily identified;
- experience elsewhere in the transparency field which highlights the importance of historical data on declarations.

The data model is a first draft: it's an opener in a conversation. Publishing its details here allows the underlying ideas to feed into further work in this domain. We hope that it might provoke interest and criticism.

The [draft data model](#) is documented in a spreadsheet for convenience. Ultimately a related data standard could be rendered in JSON schema.

 Overview of the data model



Diagram

of data model

The [draft data model](#) encapsulates an Interest object, placing it within a Declaration. This allows processes of declaration and re-declaration (or confirmation) to be treated and understood separately from the life-cycle and nature of the interest itself. For example, a person may be required to declare an interest which actually sits with a family member or business partner. And they may confirm that the same interest still exists in a future declaration.

The Interest object itself is only semi-structured. The model is built on the assumption that it is primarily **dates, people and organisations** associated with interests that people wish to link to other information. We can imagine that there might be other notable information types - for instance, location. Beginning with the most obvious and widely applicable information types seemed wise. We leave it up to data publishers to provide in their datasets the precise

field names for dates, people and organisations. More information about this feature of the model can be found below under Key features > Flexibility.

5.1 Modelling declarations

As important as the details of politicians' and prominent figures' **interests** are, so are the details of how and when those interests were **declared**. The draft model captures information about both of these things. It also allows a declaration that someone has no relevant interests to be represented.

5.2 Flexibility

In order that as many types of interests as possible might be represented by the data model, we have maintained a high level of abstraction. In practice, that means identifying the components, common to all interests, which draw most attention from data users. These, we propose, are **dates, people and organisations**. The data model allows, though, for more detail to be added: for publishers to move to a lower level of abstraction. Data profiles can be created (external to datasets) which define the specific roles of dates, people and organisations in relation to different types of interest. For example, a recommended data profile for interests in the category 'employment and earnings' might be as follows.

5.2.1 Data profile for 'Employment and earnings' interest

Object	Property	Value	Description
OrganisationField	fieldName	Organisation benefitting	The employer or organisation benefitting from the work.
OrganisationField	fieldName	Organisation paying	The organisation paying for the work if different from that benefitting from it.
DateField	fieldName	Work completion date	The date on which the work was completed.
PersonField	fieldName	Person employed	The person whose employment or earnings make this a registrable interest.

The values above would be published as part of a dataset, but their full meaning is provided by the 'description' in the data profile. Data profiles like this might be referenced from publishers' Publication Policies.

5.3 Open codelists

To aid the flexibility and applicability of the model, we have opted for open codelists (codelists from which values can be chosen or which can be ignored in favour of custom values) where possible. For example an organisation, depending on the role it has in a particular interest, may be a 'Donor, Contracting organisation, Contracted organisation, Employer, Company, Organisation, Owner'. Or it may have another role entirely. Data creators are encouraged to use the term from an open codelist that fits best, even if it is not an exact match, as this aids data interoperability.

Outstanding issues

As a draft data model, this way of representing declarations of interests is not fully formed or perfect. In particular we recognise that there is work to do on the following areas:

6.1 Publisher metadata and republishing

At the moment, the following fields relate to the acts of receiving, sourcing or publishing declarations:

- declaredTo
- dateDeclared
- datePublished
- declarationUrl

These may not adequately cover all likely use cases. For example, the following could refer to three separate organisations or bodies:

- the authority or organisation to which the declaration is made
- the original publisher of the declaration
- the republisher of the declaration

At the moment, there is only a single field (the 'declaredTo' field) which covers the first type of organisation.

Future work would ascertain likely publishing, republishing and data analysis use cases and refine related areas of the data model.

6.2 Change over time.

Work on the model proceeded with the assumption that data consumers would want to be able to pin-point dates related to people's interests (and their declaration) and track the changes to those interests, as declared, over time. Since an 'interest' is a modellable object in itself and can change over time in its nature (unlike a declaration), we suggest it is

given a unique identifier (`interestId`) and a status (`interestStatus`). In terms of how an interest relates to a declaration, the current data model allows its status to be: newly declared, updated (or confirmed), ended, or retracted.

It is envisaged that newly declared or updated interests should contain full, up-to-date details about the interest. However, an ended or retracted interest need only (as a minimum) contain the `interestID`.

This is a fairly flexible approach to handling the variety of ways that local government and other bodies publish the interests of their members, but needs to be tested in practice.

6.3 Packaging and publication format

The draft data model has a Declaration as its highest-level object. A Declaration is made by a single person (`declared-ByName`) and can contain multiple interests. Beyond that level of encapsulation we have said nothing about publishing collections or streams of declarations. The data model is amenable to various packaging and publishing options.

CHAPTER 7

About

This data model was produced by Open Data Services Co-operative as part of a project with OpenDemocracy, funded by Nesta's Future News Pilot Fund.