Defining a Model for Content Requirements from the Law: an Experience Report Appendixes

Marcello Ceci
SnT, University of Luxembourg
Luxembourg
Email: marcello.ceci@uni.lu

Domenico Bianculli
SnT, University of Luxembourg
Luxembourg
Email: domenico.bianculli@uni.lu

Lionel C. Briand[†]
Lero SFI centre for Software Research
and University of Limerick, Ireland
School of EECS, University of Ottawa, Canada
Email: lionel.briand@lero.ie

April 18, 2024

A The Design Criteria for the Content Model of Schedule A Information

The content model is shown in Fig. 1. It includes 28 direct subclasses of Information Type, some of which (e.g., Information of Units) are *generalizations* fulfilling the goal of organizing the information types for better understandability and manageability of the model.

Following the general structure of the content model (a document having a containment relation with information types organized in a taxonomy), by combining the content model with the model of fund documentation we obtain the consolidated content requirements. For example, combining the general content requirement of article 69 with point 1.9 of Schedule A ("amount of the subscribed capital with an indication of the capital paid-up"), we obtain two specific content requirements:

1. "The prospectus (or annexed management regulations or instruments of incorporation) must contain an indication of the subscribed capital", and

2. "The prospectus (or annexed management regulations or instruments of incorporation) must contain an indication of the paid-up capital".

To distinguish between the three columns of the Schedule A table (see Section **??**), we created an attribute (*reference body*) for all information types, to indicate which legal body the information is referring to. The attribute *reference_asset*, also applicable to all information types, is used to indicate which asset of the fund is being described (e.g., the whole fund, a compartment, or a class of units).

Other attributes are specific to certain information types (e.g., *membership* for the "Information on persons", *counterpart* for the "Indication of relevant contract provisions with adviser", or *aspect* for the "Information on Payment"). This information can be tied to conditions in order to retrieve and evaluate only the information that is required in the specific case. An example of the use of attributes is in point 1.8 ("names [...] of the members of the administrative body, management and supervisory boards [...]"), which translates into three content requirements:

1. "The prospectus (or annexed management regulations or instruments of incorporation) must contain an indication of the name of the persons [person name] that are members of the administrative body [membership: administrative body]",

2. "The prospectus (or annexed management regulations or instruments of incorporation) must contain an indication of the name of the persons [person name] that are members of the management board [membership: management board]", and

3. "The prospectus (or annexed management regulations or instruments of incorporation) must contain an indication of the name of the persons [person name] that are members of the supervisory board [membership: supervisory board]".
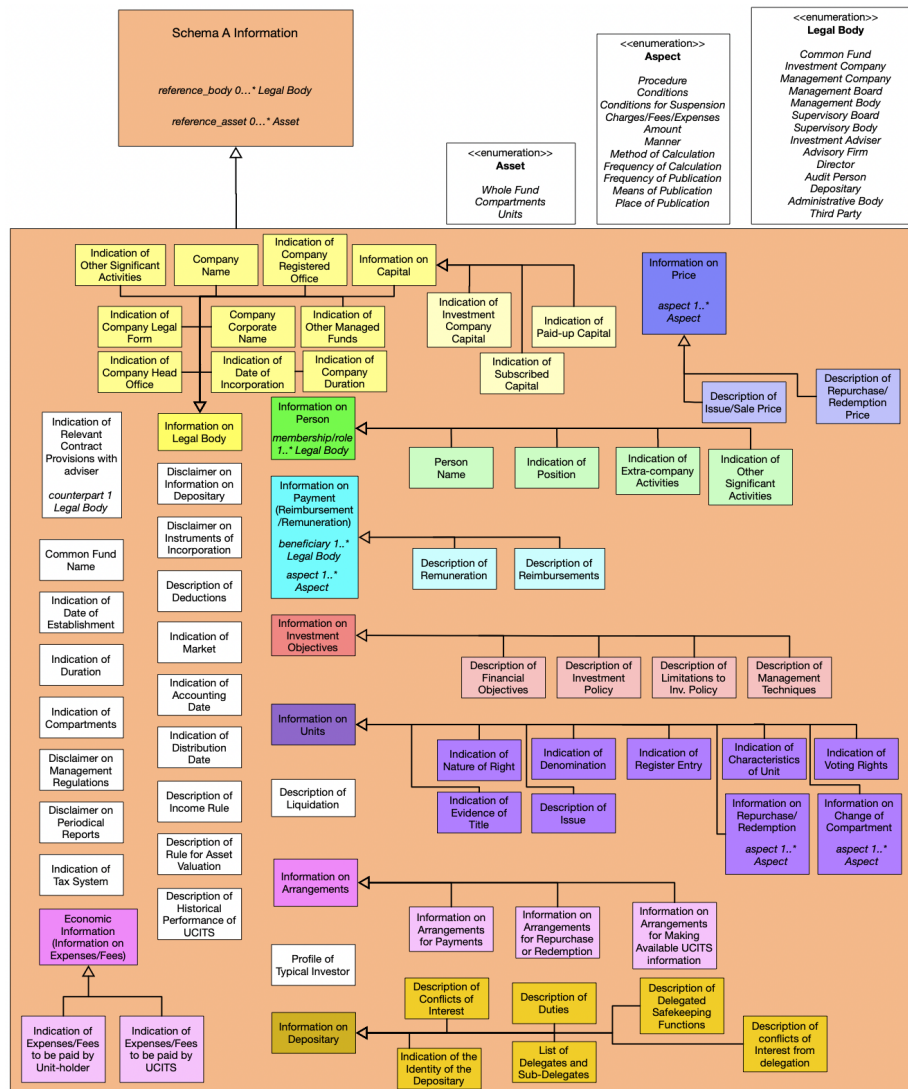
Schema A Information

reference_body 0... Legal Body*

reference_asset 0... Asset*

<<enumeration>>
Asset

Whole Fund
Compartments
Units

<<enumeration>>
Aspect

Procedure
Conditions
Conditions for Suspension
Charges/Fees/Expenses
Amount
Manner
Method of Calculation
Frequency of Calculation
Frequency of Publication
Means of Publication
Place of Publication

<<enumeration>>
Legal Body

Common Fund
Investment Company
Management Company
Management Board
Management Body
Supervisory Board
Supervisory Body
Investment Adviser
Advisory Firm
Director
Audit Person
Depositary
Administrative Body
Third Party

Indication of Other Significant Activities

Company Name

Indication of Company Registered Office

Information on Capital

Information on Price — *aspect 1..* Aspect*

Indication of Company Legal Form

Company Corporate Name

Indication of Other Managed Funds

Indication of Investment Company Capital

Indication of Paid-up Capital

Indication of Company Head Office

Indication of Date of Incorporation

Indication of Company Duration

Indication of Subscribed Capital

Description of Issue/Sale Price

Description of Repurchase/ Redemption Price

Indication of Relevant Contract Provisions with adviser — *counterpart 1 Legal Body*

Information on Legal Body

Information on Person — *membership/role 1..* Legal Body*

Person Name

Indication of Position

Indication of Extra-company Activities

Indication of Other Significant Activities

Disclaimer on Information on Depositary

Disclaimer on Instruments of Incorporation

Information on Payment (Reimbursement /Remuneration) — *beneficiary 1..* Legal Body* — *aspect 1..* Aspect*

Description of Remuneration

Description of Reimbursements

Common Fund Name

Description of Deductions

Indication of Date of Establishment

Indication of Market

Information on Investment Objectives

Description of Financial Objectives

Description of Investment Policy

Description of Limitations to Inv. Policy

Description of Management Techniques

Indication of Duration

Indication of Accounting Date

Indication of Compartments

Indication of Distribution Date

Information on Units

Indication of Nature of Right

Indication of Denomination

Indication of Register Entry

Indication of Characteristics of Unit

Indication of Voting Rights

Disclaimer on Management Regulations

Description of Income Rule

Indication of Evidence of Title

Description of Issue

Information on Repurchase/ Redemption — *aspect 1..* Aspect*

Information on Change of Compartment — *aspect 1..* Aspect*

Disclaimer on Periodical Reports

Description of Rule for Asset Valuation

Indication of Tax System

Description of Historical Performance of UCITS

Information on Arrangements

Information on Arrangements for Payments

Information on Arrangements for Repurchase or Redemption

Information on Arrangements for Making Available UCITS information

Economic Information (Information on Expenses/Fees)

Profile of Typical Investor

Description of Liquidation

Description of Conflicts of Interest

Description of Duties

Description of Delegated Safekeeping Functions

Indication of Expenses/Fees to be paid by Unit-holder

Indication of Expenses/Fees to be paid by UCITS

Information on Depositary

Indication of the Identity of the Depositary

List of Delegates and Sub-Delegates

Description of conflicts of Interest from delegation

Figure 1: The content model of Schedule A information types

A.1 Setting the granularity of representation for Schedule A provisions

The main element of complexity when representing Schedule A arises from the fact that often a Schedule A provision (sentence) expresses several information types.

Example 1. Let us consider the example of 1.17 for FCPs. It requires the prospectus of the FCP to contain a "determination of the sale or issue price and the repurchase or redemption price of units, in particular: - the method and frequency of the calculation of those prices, - information concerning the charges relating to the sale or issue and the repurchase or redemption of units, - information concerning the means, places and frequency of the publication of those prices."

Here we can initially identify three types of required information:

- The method and frequency of calculation of sale, issue, repurchase and redemption prices;

- Information on the charges relating to the sale, issue, repurchase or redemption of units;

- Information on the means, places and frequency of the publication of sale, issue, repurchase and redemption prices.

In turn, each of these can be further specified, resulting in the following list of information types:

- Description of Calculation Method for issue/sale price;

- Description of Calculation Method for repurchase/redemption price;

- Description of Calculation Frequency for issue/sale price;

- Description of Calculation Frequency for repurchase/redemption price;

- Description of Charges on issue/sale price;

- Description of Charges on repurchase/redemption price;

- Description of Means of Publication of issue/sale price;

- Description of Means of Publication of repurchase/redemption price;

- Description of Place of Publication of issue/sale price;

- Description of Place of Publication of repurchase/redemption price;

- Description of Frequency of Publication of issue/sale price;

- Description of Frequency of Publication of repurchase/redemption price.

Note that, following feedback from the supervisory authority, we have joined together "issue" and "sale" (see Section **??**), and "repurchase" and "redemption". Had we not done so, we would have twice as many information types.

Example 2. It can also happen that what initally appears as a single information type reveals, upon a deeper analysis, to be a generalization composed of several information types. Let us consider 1.12, *"Procedures and conditions of issue and/or sale of units"*. Apparently, this information type includes only two subtypes, namely procedures and conditions. When looking at the ground truth for this information type, however, it becomes clear that each of these two subtypes consists in a long span of text that may include different topics, including:

- The power to issue units;

- The power to reject subscriptions;

- The power to limit/restrict issue;

- The deadline for subscription applications (cut-off time);

- Information related to AML regulations;

- The issue time and price;

- Payment times and methods;

- The minimum subscription;

- The discipline regarding subscriptions in kind.

This poses the challenge of representing all the core information types in the content model while keeping the complexity of the model at bay.

A.2 Generalizations

We try to improve the readability of the Model by identifying multiple layers of information, i.e., *generalizations*, for some of the information types. We use generalizations also for a second purpose: to allow the specification of information types at different levels of precision/granularity. Generalizations are specified either by means of subclasses, or by means of attributes with the three literal ranges Legal Body, Asset, Aspect (see next subsection). This entails that not all information types are represented as a distinct class in the Model. In other words, there is not a 1:1 relation between the information types required by Schedule A and the classes in the Model. For example, the class "Company Corporate Name" corresponds to either 3.1 (in case the company in question is an advisory firm) or 1.1 (in case of the management company or the investment company).

We choose between subclasses or attributes using the following criteria:

- We use inheritance for the first dimension (first aspect/level of information subtype);

- We use attributes for other dimensions.

The main guidance principle for choosing which dimension is the first is to choose the dimension that is usually not totally covered by a single chunk of text. For information types that are different but can be seen as different aspects of the same information, we use attributes. For example, if "calculation method for issue price" and "calculation method for redemption price" are often found in different chunks of text, while "calculation method for issue price" and "calculation frequency of issue price" are often found in the same chunk of text, then we will consider "issue/redemption" as the first dimension (and use inheritance to represent it), and we will consider "calculation method/frequency" as the second dimension (and use an attribute to represent it). In this way, we limit the duplication of information when instantiating the model.

A.3   Literals Enumerations

Literals used to specify the values of attributes are grouped into three enumerations:

- "Asset" includes the assets of the fund: the fund itself, the units classes, and the compartments. Because sometimes a certain information is required for each of these assets, we added a generic attribute "reference asset" that can be used to specify each information.

- "Legal Body" includes all legal subjects that are referenced in Schedule A. We use this class to specify certain generalizations (e.g., "company name") into the single information types that are required by Schedule A. We also use this class to distinguish between the three columns of the Schedule A table.

- For "Aspect" the semantics is not as strictly defined: it is rather a "residual" category including all entities that do not represent a legal body or an asset of the fund. This enumeration contains abstract concepts, that could not be represented as information types according to our criteria for classification. Following feedback from the supervisory authority, we put fees, charges, and expenses, together as a single literal (since they are considered as interchangeable terms).

A.4   Naming convention for information types

For labelling information types we followed a naming convention:

(a) We add no prefix to information types that correspond to names or labels (e.g. Fund Name);

(b) We add the prefix "Disclaimer on..." for information types that express fixed content (e.g., Disclaimer on Management Regulations);

(c) We add the prefix "Indication of..." for information types that contain a specific info in form of a string (e.g., Indication of Duration);

(d) We add the prefix "Description of..." for information types that describe (summarize) information (e.g., Description of Remuneration);

(e) We add the prefix "Information on. . . " for generic information types that contain different subtypes (e.g., Information on Legal Body).

Initially we proposed a different definition for "Information on. . . ", identifying strings that "summarize information about an entity or value". The partner supervisory authority rejected this proposal, noting that it was too similar to "Description of. . . ". We therefore repurposed "Information on. . . " as a generic indicator. Note that in Schedule A we find "Information on. . . ", but we treat these information types as "Description of. . . ".

B   The Classes representing Schedule A Information

Following the criteria indicated in the previous sections results in the definition of 28 subconcepts of Schedule A Information, representing either core information types or their first dimension of generalization (see A.2):

1. Information on Person

2. Information on Price

3. Information on Payment (Reimbursement/Remuneration)

4. Information on Legal Body

5. Information on Investment Objectives

6. Information on Units

7. Information on Arrangements

8. Information on Depositary

9. Economic Information (Information on Expenses/Fees)

10. Description of Liquidation

11. Indication of Relevant Contract Provisions with Adviser

12. Common Fund Name

13. Indication of Date of Establishment

14. Indication of Duration

15. Indication of Compartments

16. Indication of Tax System

17. Indication of Market

18. Indication of Accounting Date

19. Indication of Distribution Date

20. Disclaimer on Management Regulations

21. Disclaimer on Periodical Reports

22. Disclaimer on Information on Depositary

23. Disclaimer on Instruments of Incorporation

24. Description of Deductions

25. Description of Income Rule

26. Description of Rule for Asset Valuation

27. Historical Performance of UCITS

28. Profile of Typical Investor

While classes 12–28 represent actual information types, classes 1–11 represent generalizations. Of those, classes 4–9 express multiple information by way of subclasses, classes 10–11 express multiple information types by way of attributes, and classes 1–3 use both subclasses and attributes. See Figures 2 and 3 for the mapping between the classes of the model and the points of Schedule A.

Following is a description of each generalization:

1. *Information on Person*: it groups together all information on natural persons;

2. *Information on Price*: it groups together all information on the price of units;

3. *Information on Payment (Reimbursement/Remuneration)*: it groups together all information on payments performed by the Fund to other natural or moral persons;

4. *Information on Legal Body*: it groups together all information on moral persons;

5. *Information on Investment Objectives*: it groups together all information on investment objectives as described in point 1.15 of Schedule A;

6. *Information on Units*: it groups together all information on units (except for the information on price of units);

7. *Information on Arrangements*: it groups together all information on arrangements made by the Fund, as described in point 4 of Schedule A;

8. *Information on Depositary*: it groups together all information on the depositary (note that this information should be in "Information on Legal Body" but we kept it distinct since there are no information types in common among the two classes; note also that according to CSSF the "Identity of the Depositary" corresponds to the name and address);

9. *Economic Information (Information on Expenses/Fees)*: it groups together all information on expenses and fees as described in point 6 of Schedule A;

10. *Information on Liquidation*: it groups together all information on the Liquidation of funds, sub-funds, or units classes (according to the supervisory authority, the liquidation of units classes should be treated as liquidation of sub-funds);

11. *Indication of Relevant Contract Provisions with Adviser*: it groups together all information on relevant contract provisions valid between with the adviser and either the management company or the investment company, as described in point 3.2 of Schedule A.

Following is a list of the information types that map to multiple parts of Schedule A with an indication of how attributes determine the ultimate classification:
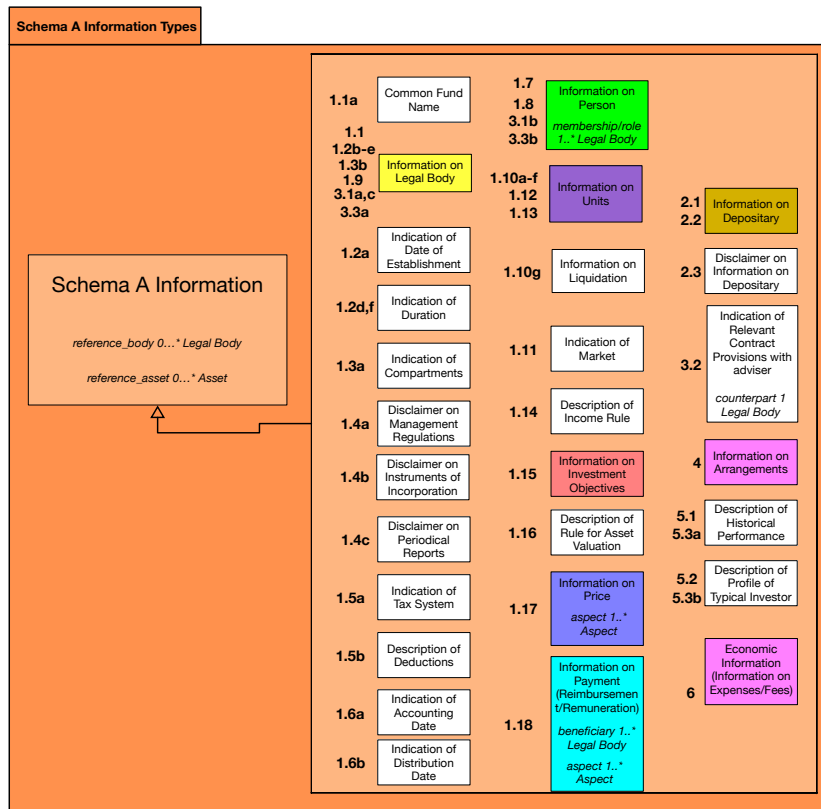
Figure 2: The mapping of the information types in our Model to the corresponding sections in Schedule A
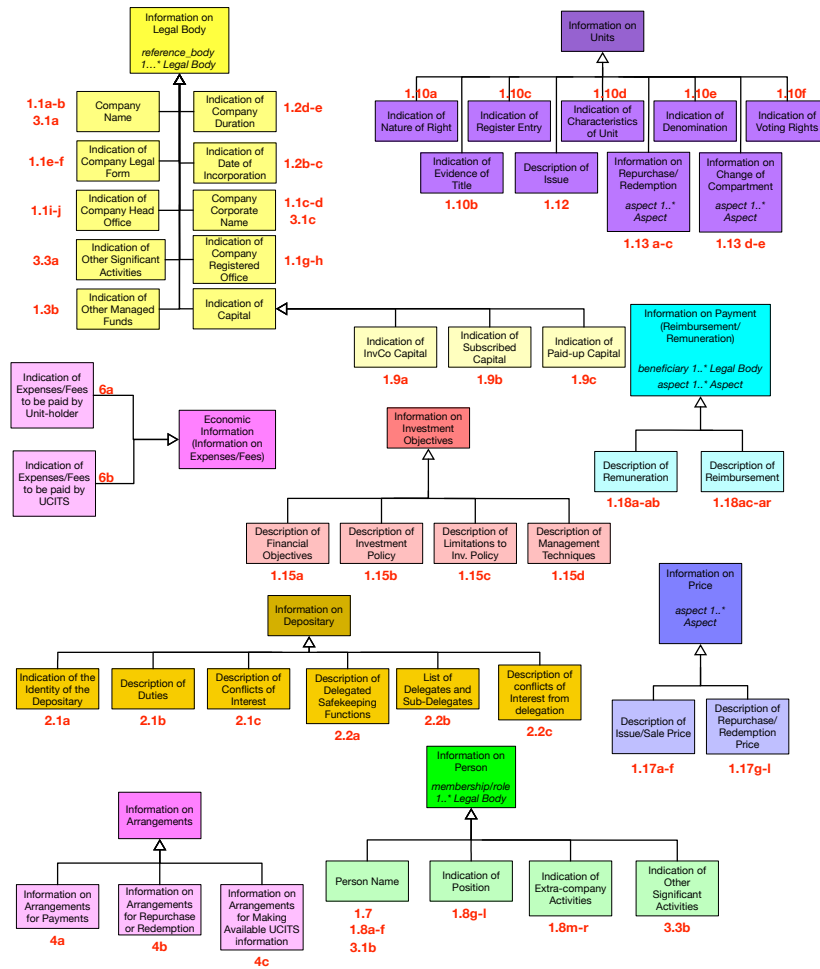
Figure 3: The mapping of the information sub-types in our Model to the corresponding sections in Schedule A

(1) *Information on Liquidation*: we distinguish according to the range of *reference asset*. If "whole fund", then the information corresponds to 1.10g or 1.10i (see below); if "compartments" or "units class" then the information corresponds to 1.10i or 1.10j (see below). We also distinguish according to the range of *aspect*. If "conditions", then the information corresponds to 1.10g (for whole fund) or 1.10i (compartments or units classes); if "procedure", then the information corresponds to 1.10h (for whole fund) or 1.10j (compartments or units classes).

(2) *Information on Issue/Sale*: we distinguish according to the range of *aspect*. If "conditions", then the information corresponds to 1.12a; if "procedure", then the information corresponds to 1.12b.

(3) *Information on Repurchase/Redemption*: we distinguish according to the range of *aspect*. If "conditions", then the information corresponds to 1.13a; if "procedure", then the information corresponds to 1.13b; if "suspension conditions", then the information corresponds to 1.13c.

(4) *Information on Change of Compartment*: we distinguish according to the range of *aspect*. If "procedure", then the information corresponds to 1.13d; if "fees", then the information corresponds to 1.13e.

(5) *Company Name*: we distinguish according to the range of *reference body*. If "Investment Company", then the information corresponds to 1.1a; if "Management Company", then the information corresponds to 1.1b; if "Advisory Firm", then the information corresponds to 3.1a.

(6) *Company Corporate Name*: we distinguish according to the range of *reference body*. If "Investment Company", then the information corresponds to 1.1c; if "Management Company", then the information corresponds to 1.1d; if "Advisory Firm", then the information corresponds to 3.1b.

(7) *Indication of Company Legal Form*: we distinguish according to the range of *reference body*. If "Investment Company", then the information corresponds to 1.1e; if "Management Company", then the information corresponds to 1.1f.

(8) *Indication of Registered Office*: we distinguish according to the range of *reference body*. If "Investment Company", then the information corresponds to 1.1g; if "Management Company", then the information corresponds to 1.1h.

(9) *Indication of Head Office*: we distinguish according to the range of *reference body*. If "Investment Company", then the information corresponds to 1.1i; if "Management Company", then the information corresponds to 1.1j.

(10) *Indication of Date of Incorporation*: we distinguish according to the range of *reference body*. If "Investment Company", then the information corresponds to 1.2b; if "Management Company", then the information corresponds to 1.2c.

(11) *Indication of Duration*: we distinguish according to the range of *reference body* or *asset*. If "Investment Company", then the information corresponds to 1.2d; if

"Management Company", then the information corresponds to 1.2e; if "sub-funds", then the information corresponds to 1.2f.

(12) *Description of Remuneration*: we distinguish according to the range of *reference body*, *beneficiary* and *aspect*. Regarding *reference body*: if "Investment Company", then the information corresponds to 1.18a–h, m–ab; if "Common Fund", then the information corresponds to 1.18i–l, u–ab. Regarding *beneficiary*: if "Directors", then the information corresponds to 1.18a–d; if "Administrative Body", then the information corresponds to 1.18e–h; if "Management Company", then the information corresponds to 1.18i–p; if "Supervisory Body", then the information corresponds to 1.18q–t; if "Depositary", then the information corresponds to 1.18u–x; if "Third Party", then the information corresponds to 1.18y–ab. Regarding *aspect*: If "Calculation Method", then the information corresponds to 1.18a/e/i/m/q/u/y; if "Calculation Frequency", then the information corresponds to 1.18b/f/j/n/r/v/z; if "Amount", then the information corresponds to 1.18c/g/k/o/s/w/aa; if "Manner", then the information corresponds to 1.18d/h/l/p/t/x/ab.

(13) *Description of Reimbursement*: we distinguish according to the range of *reference body*, *beneficiary* and *aspect*. Regarding *reference body*: if "Investment Company", then the information corresponds to 1.18ac–af, ak–ar; if "Common Fund", then the information corresponds to 1.18ag–ar. Regarding *beneficiary*: if "Directors", then the information corresponds to 1.18ac–af; if "Management Company", then the information corresponds to 1.18ag–aj; if "Depositary", then the information corresponds to 1.18ak–an; if "Third Party", then the information corresponds to 1.18ao–ar. Regarding *aspect*: If "Calculation Method", then the information corresponds to 1.18ac/ag/ak/ao; if "Calculation Frequency", then the information corresponds to 1.18ad/ah/al/ap; if "Amount", then the information corresponds to 1.18ae/ai/am/aq; if "Manner", then the information corresponds to 1.18af/aj/an/ar.

(14) *Description of Issue/Sale Price*: we distinguish according to the range of *aspect*. If "Calculation Method", then the information corresponds to 1.17a; if "Calculation Frequency", then the information corresponds to 1.17b; if "Charges", then the information corresponds to 1.17c; if "Means of Publication", then the information corresponds to 1.17d; if "Place of Publication", then the information corresponds to 1.17e; if "Frequency of Publication", then the information corresponds to 1.17f.

(15) *Description of Repurchase/Redemption Price*: we distinguish according to the range of *aspect*. If "Calculation Method", then the information corresponds to 1.17g; if "Calculation Frequency", then the information corresponds to 1.17h; if "Charges", then the information corresponds to 1.17i; if "Means of Publication", then the information corresponds to 1.17j; if "Place of Publication", then the information corresponds to 1.17k; if "Frequency of Publication", then the information corresponds to 1.17l.

(16) *Person Name*: we distinguish according to the range of *reference body* and *membership/role*. Regarding *reference body*: if "Investment Company", then the information corresponds to 1.8a–c; if "Management Company", then the information corresponds to 1.8d–f. Regarding *membership/role*: if "Audit Person", then the

information corresponds to 1.7; if "Administrative Body", then the information corresponds to 1.8a or 1.8d; if "Management Body", then the information corresponds to 1.8b; if "Supervisory Body", then the information corresponds to 1.8c; if "Management Board", then the information corresponds to 1.8e; if "Supervisory Board", then the information corresponds to 1.8f; if "Investment Adviser", then the information corresponds to 3.1b.

(17) *Indication of Position*: we distinguish according to the range of *reference body* and *membership/role*. Regarding *reference body*: if "Investment Company", then the information corresponds to 1.8g–i; if "Management Company", then the information corresponds to 1.8j–l. Regarding *membership/role*: if "Administrative Body", then the information corresponds to 1.8g or 1.8j; if "Management Body", then the information corresponds to 1.8h; if "Supervisory Body", then the information corresponds to 1.8i; if "Management Board", then the information corresponds to 1.8k; if "Supervisory Board", then the information corresponds to 1.8l.

(18) *Indication of Extra Company Activities*: we distinguish according to the range of *reference body* and *membership/role*. Regarding *reference body*: If "Investment Company", then the information corresponds to 1.8m–o; if "Management Company", then the information corresponds to 1.8p–r. Regarding *membership/role*: if "Administrative Body", then the information corresponds to 1.8m or 1.8p; if "Management Body", then the information corresponds to 1.8n; if "Supervisory Body", then the information corresponds to 1.8o; if "Management Board", then the information corresponds to 1.8q; if "Supervisory Board", then the information corresponds to 1.8r.

(19) *Indication of Relevant Contract Provisions with Adviser*: we distinguish according to the range of *counterpart*. if "Management Company", then the information corresponds to 3.2a; if "Investment Company", then the information corresponds to 3.2b.

(20) *Indication of Historical Performance*: we distinguish according to the range of *reference asset*. if "Whole Fund", then the information corresponds to 5.1; if "Sub-Funds", then the information corresponds to 5.3a.

(21) *Indication of Profile of Typical Investor*: we distinguish according to the range of *reference asset*. if "Whole Fund", then the information corresponds to 5.2; if "Sub-Funds", then the information corresponds to 5.3b.

Containment vs Subsumption. We decided to not use containment even in cases when it could be used ("Information on Investment Objectives"). It could be argued that Information on Investment Objectives is a composition of its subtypes (Description of Financial Objectives, Description of Investment Policy, Description of Limitations to Investment Policy, Description of Management Techniques). However, considering that the semantics of the Model focus on finding chunks of text, we preferred to have the same semantics for all information types, i.e., we treat all "first layer classes" as either information types or generalizations. Investment Objectives are therefore treated as a generalization rather than a composition in our Model.

C   The Domain Model of Fund-related Entities

The content model presented in A supports query answering on Schedule A information. In order to support automated compliance checking, further modeling is needed. It is in fact necessary to include concepts that can be used to express pre-conditions or constraints. For example, the information of 1.3 ("indication of investment compartments") is only required if the fund has different investment compartments. To be able to automatically check this circumstance, the content model must be extended with a domain model on fund-related entities, which will in turn include the concepts of "fund described by the fund documentation" and "investment compartment" that the fund may or may not have (see Figure **??**). Note that these additional concepts may be expressed in spans of text that have been considered non-core (or not been considered at all) during the model refinement phase (see Section **??**).

The complete model of fund-related entities is shown in Figure 4. In the graph, the darker boxes represent UCITS funds and its subcategories, dark boxes represent elements of UCITS funds, and transparent boxes represent (natural and legal) persons.

Below there is a list of the entities and relations in the model:

- UCITS Fund (*appoints_depositary* 1+ depositary, *appoints_manager* 0+ Investment Manager, *has_compartment* 0+ Compartment, *issues* 1+ Unit, *appoints_auditor* 0+ Auditor, *appoints_distributor* 0+ Distributor):

    - FCP

    - SICAV (*appoints_bod* 1 Board of Directors)
        * Self-managed SICAV (*appoints_management_board* 0+ Management Board)
        * SICAV which designates ManCo (*appoints_manco* 1+ Management Company)

- Compartment

- Unit

- Unit Class (*describes* 0+ Unit)

- Unitholder (*owns* 1+ Unit)

- Depositary

- Investment Manager (*manages_compartment* 0+ Compartment)

- Auditor

- Administrative Body

    - Board of Directors (*represents* 1 SICAV, *represents* 1 Management Company, *has_member* 1+ Director)
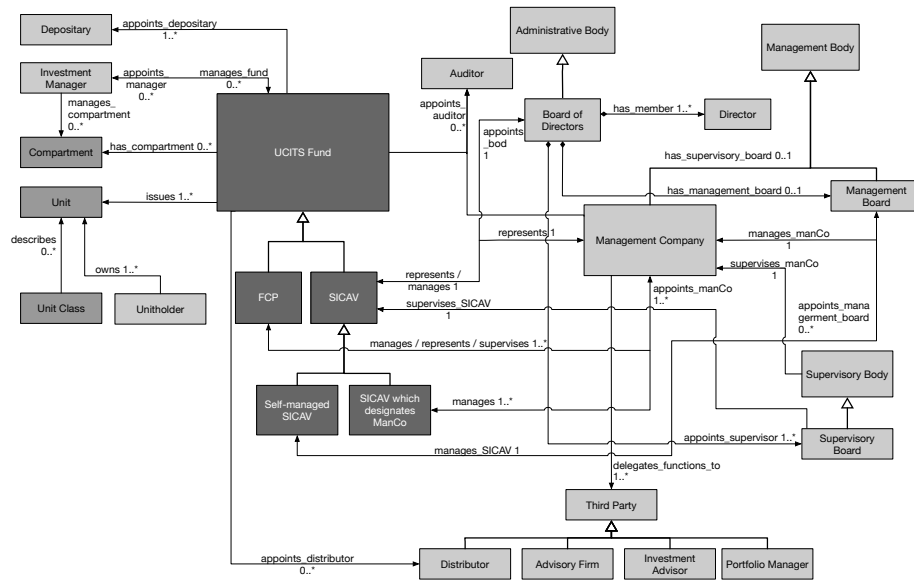
- Director

Figure 4: The model for fund-related entities

- Management Body

    - Management Board (*manages_manco* 1 Management Company, *manages_SICAV* 1 Self-managed SICAV)

    - Management Company (*manages* 1 SICAV, *delegates_functions_to* 1+ Third Party, *manages* 1+ SICAV which designates ManCo)

- Supervisoy Body (*supervises_manco* 1 Management Company)

    - Supervisory Board (*supervises_sicav* 1 SICAV)

- Third Party

    - Distributor

    - Advisory Firm

    - Investment Advisor

    - Portfolio Manager