FORWARD

Project Number:	**325135**
Project Acronym:	**FORWARD**
Project Title:	**Framework for a EU-wide Audiovisual Orphan Works Registry**
Instrument:	Pilot B
Thematic Priority:	CIP-ICT-PSP-2012-6

D4.2 Definition of System Workflow

Due Date:	31/10/2014
Submission Date:	….
Start Date of Project:	01/07/2013
Duration of Project:	36 months
Partner in Charge of Deliverable	CINECA, EYE, DFI
Version Status	FINAL
Dissemination Level	PU[1]
File Name:	DELIVERABLE 4.2

[1] Check List of Deliverables, pp. 2-4 of DoW

Revision History

Revision	Date	Author	Organisation	Description
N°1	22/09/2014	CINECA	CINECA	First analysis of the decision tree and algorithms
N°2	09/10/2014	CINECA	CINECA	FORWARD workflow completed and ready for internal assessment
N°3	07/10/2014	EYE	EYE	First Draft of the metadata needed to FORWARD system and ERD
N°4	17/10/2014	DFI	DFI	First Draft of the Metadata Enrichment requirements
N°5	22/10/2014	CINECA	CINECA	Draft – workflow reviewed, Added ARROW vs FORWARD
N°6	24/10/2014	CINECA, EYE, DFI	CINECA	D4.2 Integrated version and revisions
N°7	28/10/2014	CINECA, EYE, KAVI	CINECA, EYE, KAVI	Revisions
N°8	29/10/2014	DFI, DIF, EYE	DFI, DIF, EYE	Workflow Configuration parameters sheet added in Annexes
N° 9	8/11/2014	NMazzanti	CRB	Revisions
N° 10	11/11/2014	GScipione	CINECA	FINAL

Table of Contents

EXECUTIVE SUMMARY

The initial objectives of task 4.2 were multiple; in the DoW were combined as follow:

- analysis and design of expected workflow for assessing rights status(es) of AV works,
- query the system by users and redirect to appropriate resource;
- design tools for enabling metadata enrichment with rights data and analysis of the existing metadata structures.

In this deliverable the activities are analysed and the main outcomes presented in order to provide a comprehensive description of the results achieved. .

These results can be summarised as follows:

FORWARD workflow:

A detailed analysis of the FORWARD system workflow and in particular of the steps and processes necessary to assess the rights status of a work was presented. After analysing the first decision trees designed on the basis of input from The Netherlands, Germany and Finland, CINECA broke each workflow into more granular and generic elements in order to identify logical processes, components and variations between the different countries.

The FORWARD workflow maintains the three macro-steps identified in D4.1, here named as follows:

- Upload Metadata,
- Establish Copyright Status,
- Rightsholder Determination -Establish Orphan Status.

For the sake of clarity it was decided to distinguish in each macro step the logical decision tree (algorithm) from the effective workflow that FORWARD will employ. Each macro-step is mainly composed by two sections:

- The first describes the decision logic algorithms built on the basis of the provided or gathered information;
- the second describes the related workflow by specifying the way the information is retrieved and processed

In order to manage the variations of national law and operating procedure of each country, the FORWARD system was designed to be configurable and a workflow configuration parameters list was implemented.

Metadata Enrichment requirements:

This section presents the analysis and definition of the existing metadata structures required for the execution of the FORWARD workflow. Based on this output, three enrichment phases were identified within the project. The first is named “FHI local enrichment” and implies that FHI will enrich their local informational resources in order to improve/complete the available information. The other two enrichment phases, respectively named “FW enrichment phase for ADiSP” and “FW enrichment phase for CoSAP”, include the enrichment required for the completion of the ADiSP and CoSAP/Orphan status processes (as defined in D3.2).

In chapter 2 we identified the required metadata for the FORWARD system, this can be divided in cinematographic information (required for the identification of an AV-work, such as title, main contributors, country of origin, etc.) and Web Identity Search and company information (both needed to identify rightsholders, such as contact information, successors, etc.). We also described the field types of the FORWARD system required metadata and the current database structure of the FHI’s (previously analysed in D4.1). The

Entity Relationship Model (ERD) is a logical follow-up of the above inventory. It clarifies the structure of the metadata, which consists of three sections: filmographic information, diligent search and tree configuration.

The analysis of the metadata required by the FORWARD workflow constitutes an initial base for the definition of the appropriate metadata schema for AV rights information, that in DoW was foreseen within task 4.2. As the design of the metadata schema depends on such information, it was decided to postpone this section at the end of the design phase (task 4.3), as through D4.2 it was possible to identify the metadata elements from Denmark, Finland and the Netherlands, the three countries where the OWD was implemented (or closer to implementation).

After completion of all Decision Trees by FORWARD partners, we will have a clearer picture of all the metadata elements needed by FORWARD system. It is therefore logical to compile the metadata schema after having received the input from all our partners, planned for the end of the design phase.

PURPOSE AND STRUCTURE OF THE DOCUMENT

This document presents the description of the FORWARD workflow that was defined based on the initial requirements described in D3.2 and on input received during the definition of the legal decision trees. Moreover, it gives a general overview of the metadata that are necessary for the workflow to assess the Copyright and Orphan status of the AV works and gives an overview of the phases foreseen for the metadata enrichment requirements.

The document is composed of 3 main chapters.

- Chapter 1 contains a detailed analysis of the FORWARD system workflow and in particular the algorithms and the steps and processes that are necessary to assess the rights status of a work.
- Chapter 2 describes the metadata that are necessary for the FORWARD system and the type and structure of the fields. It also contains the different phases of the metadata enrichment requirements.
- Chapter 3 outlines the main conclusions reached during the course of the task 4.2 and the next steps.

RELATION WITH OTHER DOCUMENTS

The definition of the FORWARD system workflow is based on previous deliverables:

- "D3.1 Sector analysis and user requirements" delivered by EYE.
- "D3.2 System requirements" delivered by CRB
- "D4.1 Analysis of Available Resources in the AV sector" delivered by EYE.

D3.1 laid the ground for the determination of information resources available within the partnership and outside, including mechanisms for rights status assessment in national environments. D4.1 provided an analysis of the legal groundwork required for the workflow design necessary to assess the rights status in Task 4.2.

Chapter 5 of D4.1 was also an input for the Metadata enrichment analysis (Chapter 2), as it provided a detailed analysis of the available systems and data of the partners' resources.

Additionally, the deliverable D3.2 "System requirements" delivered at the end of M16 is strictly linked to this document as it contains a high-level description of system requirements and functionalities.

1. FORWARD WORKFLOW

This section details the FORWARD system workflow and in particular the steps and processes necessary to assess the rights status of a work.

The legal analysis performed in deliverables D3.1 and D4.1 constitutes the basis for the definition of the FORWARD workflow and its algorithms.

For the sake of modularity, configurability, flexibility, extensibility and re-usability, after analysing the first decision tree designed based on the input from The Netherlands, Germany and Finland, CINECA proceeded to break each workflow into more granular and generic elements in order to identify the logical processes, components and variations between the different countries.

Then a generic FORWARD workflow was defined, composed by all the components and elements analysed so far. In order to reach consensus regarding the outline and structure of the workflow, the Dutch, German and Finnish Legal Decision Trees were considered.

In order to manage the variations of national law and operating procedure of each country, the FORWARD system will define a workflow instance for every country. Such instances will be set up based on the workflow configuration settings provided for each country by the partner in charge.

The workflow can be configured at two different levels:

1. Workflow branches to be applied or not
2. Parameters for each selected branch to be applied or not

The present workflow was approved by Dutch, German and Finnish archives, where a national Orphan Works Law was already implemented or is close to be. For the other countries, it was agreed that the configuration of their workflow will be done during the course of the project, following the adoption of OW legislation in each country.

The FORWARD workflow is broken into three macro-step procedures as depicted in the following figure:

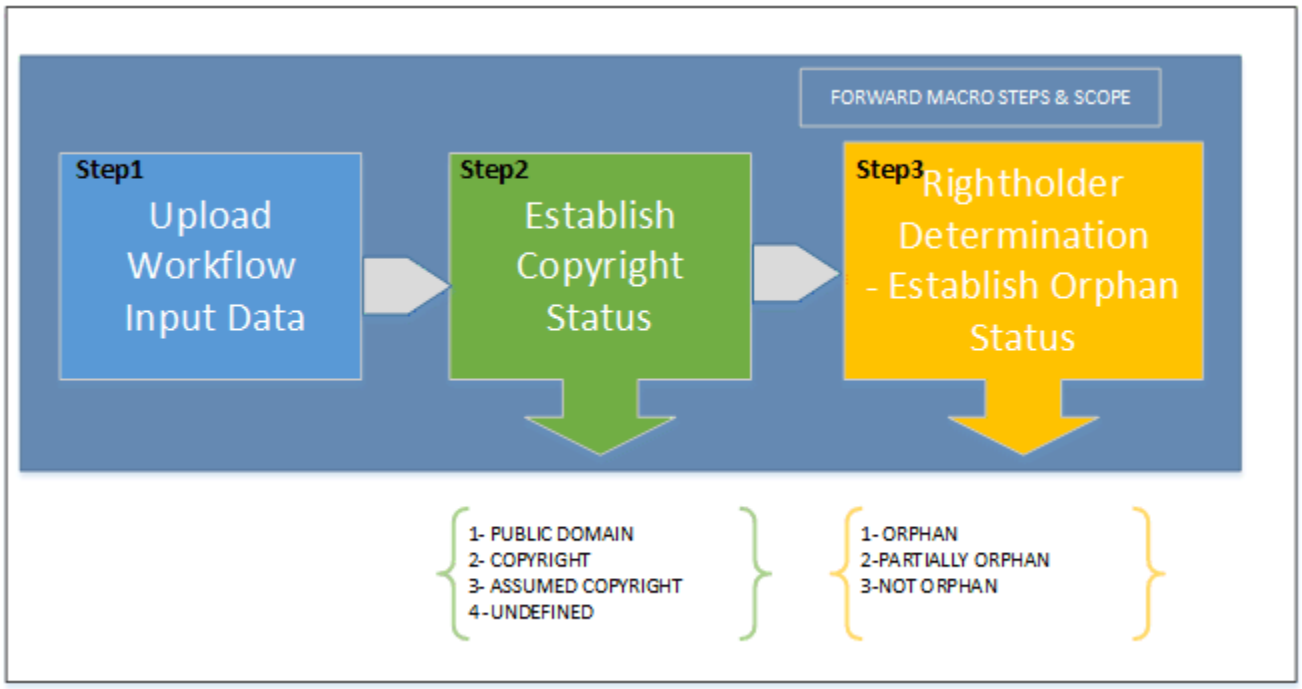


Figure 1 - FORWARD Workflow Macro Steps

- **Upload Workflow Data** – it is the starting point of the workflow. It enables users to upload their request(s) and establish the national workflow country to be applied.
- **Establish Copyright Status** – the aim of this step is to establish the copyright status of a work. As output of this step the work can be given one of the following statuses: "PUBLIC DOMAIN","UNDER COPYRIGHT", "ASSUMED COPYRIGHT", "UNDEFINED".
- **Establish Orphan Status** – the aim of this step is to determine if a work with the status of COPYRIGHTED or ASSUMED COPYRIGHT[ED] is ORPHAN, PARTIALLY ORPHAN or NOT ORPHAN. This step is thus carried out only in the case where the works copyright status is "COPYRIGHT" or "ASSUMED COPYRIGHT".

Although each macro step will be described in the following sections, before proceeding it is important to underline the top-down methodology used to define the complete FORWARD workflow.

a) First the legal decision trees of the above mentioned countries were analysed (D3.1 and D4.1)
b) Next a more complex decision tree that handles and matches all the above decision trees was defined
c) These logical decision trees (algorithms) generated the first base for the modelling of the FORWARD system workflow. In order to fully allow assessment of the rights status of an AV work the following aspects was taken into consideration in the definition of the FORWARD workflow:
   - National variations in terms of law and operating procedures
   - Missing information in the metadata describing AV works for which a diligent search is needed.
   - Available Resource Analysis
   - Diligent search tracking
   - Flexibility and Extensibility of the workflow

1.1 Upload Workflow Input Data

The following figure describes the first macro-step of the FORWARD Workflow.

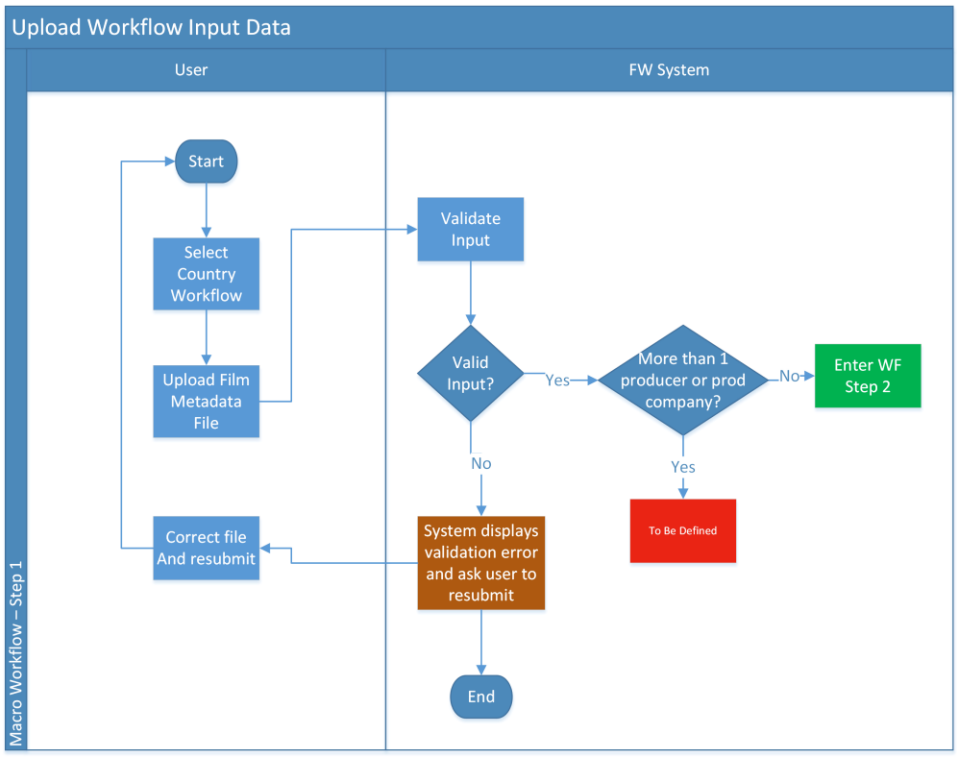


Figure 2 – Upload Workflow Input Data

As depicted by the above swim lane diagram, in order to start the FORWARD Workflow, the user must provide the system with two types of mandatory information:

- A file containing the AV work metadata according to a metadata format established by FORWARD – i.e. the AV work metadata file is prepared beforehand by the user.
- The national workflow to be applied – The national workflow to be applied is not deduced by the AV work metadata or by the user profile. This design provides higher flexibility since it allows users to perform the diligent search in a country that is different from that of the Beneficiary User group to which they belong.

Next, the FORWARD system verifies that the input file is valid according to the FORWARD established metadata schema. It also checks whether the selected national workflow is configured and can be applied. The national workflow country is specified prior to file validation in order to allow the system to customise the input file validation process by nation/country if necessary.

In case of successful input validation, the system checks whether more than one producer or production companies are present in the submitted file. Only if there is just one producer or production company the workflow is ready to start the second macro step (Establish Copyright Status). In the other case (multiple producers) the system behaviour still needs to be defined and can be modelled later on during the project.

In case of input validation errors, the system displays to the user all the failure reasons and invites him to correct

and resubmit the file.

Please note that the request upload process described hereby is just a preliminary approach and may be subject to variation during the FORWARD system design and implementation phases.

1.2 Establish Copyright Status

Before describing the FORWARD System Copyright Status workflow, let's briefly describe the logical decision tree related with establishing the copyright status of the AV work. Please note that the logical decision tree describes the decision logic taken based on the provided or gathered information, and it does not describe the way this information is retrieved. The way the information is retrieved is described in the related workflow.

In the current stage we are not considering contributors belonging to countries that are not in the EEA (European Economic Area).

1.2.1 Copyright Status Algorithm

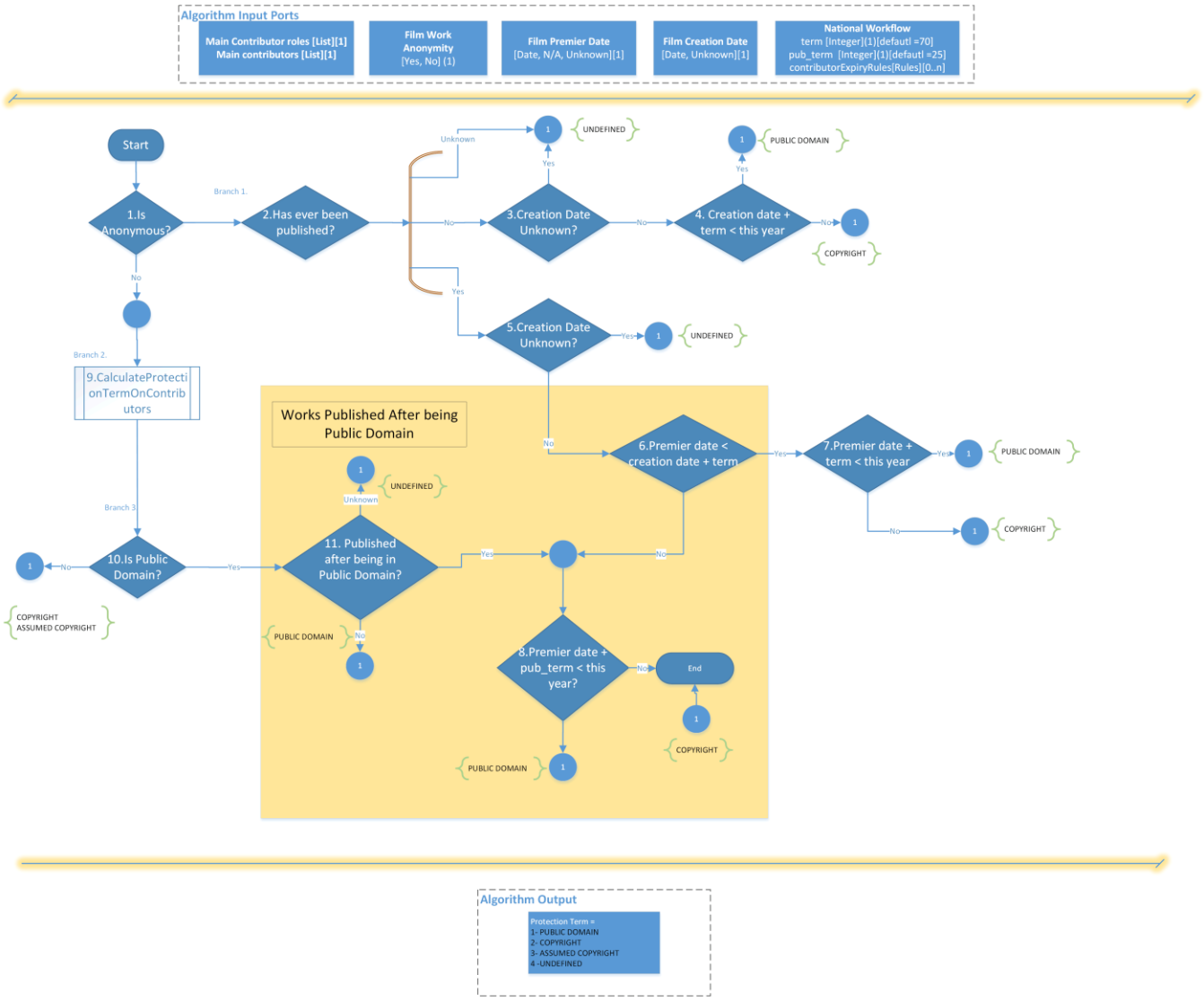


Figure 3 - Copyright Status Algorithm – (1/2)

The above figure describes the copyright status algorithm in terms of inputs, decision logic and outputs.

Please note that in all the visuals, whenever a decision or process ends on an arrow indicating a circle with a number, it means that the process enters the state indicated by such number. For example in the above case, all entities that have an outgoing arrow to the circle labelled 1 enter the End state. This is just a flow diagram convention employed in order to avoid “overwhelming” connector usage and to increase diagram clarity.

Algorithm inputs are indicated in the Algorithm Input Ports area. The inputs are of two categories:

- AV work work related metadata
- Workflow related metadata

The former category concerns information that is specific to the AV work such as film anonymity, Main

contributor roles, Main Role Contributors, AV work premiere date, AV work creation date which are provided by the user.

The latter category concerns information that is specific to the national country workflow being applied which is pre-configured in the FORWARD system.

For each input, the expected type, cardinality and default values are specified in the figure.

Algorithm outputs are obtained as a result of applying the decision logic to the provided inputs. In this case the algorithm output consists in the determination of the copyright status of a work and can have the values indicated in Algorithm Output area at the end of the figure. For the sake of clarity, in the above figure the algorithm output is specified in curly brackets.

Algorithm logic – The first step of the algorithm determines if the AV work is anonymous or not. Please note that during the different exchanges we had with the partners, there were some "differences" among them regarding the definition of anonymous work (See Glossary). For this reason the FORWARD workflow will not establish by itself if the work is anonymous or not but the information whether a work is anonymous or not will be provided to the FORWARD system explicitly by the users. It's up to the user to apply the **correct** definition of anonymous based on the country workflow that is being considered.

In an affirmative case, branch 1 of the algorithm is entered and the copyright status of the AV work depends on whether the AV work has ever been published or not.

a) In case the AV work **has not been published** (*No arrow, which corresponds also to N/A input value*), the creation date is used to establish the copyright status.
- In case the creation date is known, the **creation date** and the specific national country workflow **term** are used to establish the copyright status. The AV work is considered to be in the PUBLIC DOMAIN only if it has been created more than a number of years ago; this number of years corresponds to the specific national country workflow **term** parameter.
- In case the creation date is Unknown (never published anonymous work for which no creation date is known) the algorithm yields UNDEFINED.

b) In case the AV work **was published** (*Yes arro*w), the algorithm tries to compare the **premiere date** with the **creation date** (diamond numbered 6 in the above figure) in order to establish whether the AV work was published while in the PUBLIC DOMAIN (more than **term** years after the creation date)
- If the AV work was not published while in the public domain (Yes arrow out of Step.6), then the AV work is considered to be in the PUBLIC DOMAIN only if it was published more than a number of years ago (this number of years corresponds to the specific national country workflow **term** parameter); in the other case, the AV work is considered to be under COPYRIGHT.
- If the AV work was published while it was in the public domain, then the **premiere_date** and the specific national country workflow **pub_term** are used to establish the copyright status. In this case, the AV work is considered to be in the PUBLIC DOMAIN only if it has been published more than a number of years ago (this number of years corresponds to the specific national country workflow **pub_term** parameter). In the other case the AV work is considered to be under COPYRIGHT.

In case the AV work has been published, but the creation date is not known then the system cannot deduce the copyright status and the UNDEFINED value is returned.

Further discussions with the FHI need to take place in order to decide for each of them the copyright status on anonymous works needs to be managed. From the exchanges had so far with the partners, the following differences have been identified:.

- Some partners want to consider just the creation date (so ignore whether an anonymous work was

ever published or not – avoiding steps 5, 6, 7, and 8)
- Others want to consider also whether a work has been published or not. If the work has been published, some of the partners want to evaluate further if it was published while in public domain and others do not.

In the above figure the generic case of the anonymous work algorithm has been provided and customisations can be performed by each country during the workflow definition. See section 1.2.2.3.

c) In case a premiere date is not found (*Unknown arrow*) and it cannot be established whether the AV work has ever been published the algorithm cannot determine the protection term and thus the UNDEFINED values is returned.

If the AV work work is not anonymous then the second branch is taken into consideration and the establishing of the copyright status depends on the information related with the AV work work authors and the specific national country workflow. The algorithm is described in the next figure labelled "Copyright Status Algorithm (2/2)".

As a result of this sub-algorithm one of the following statuses is obtained: PUBLIC DOMAIN; COPYRIGHT, ASSUMED COPYRIGHT. In case the sub-algorithm result is COPYRIGHT or ASSUMED COPYRIGHT the algorithm ends, otherwise branch 3 is taken.

The aim of branch 3 is to establish whether the AV work has only been published after the work fell into the public domain and if yes, checks whether copyright related with the publication still exist:

a) In case the AV work has been published (*Yes arro*w), the **premiere date** and the specific national country workflow **pub_term** are used to establish the copyright status. In this case, the AV work is considered to be in the PUBLIC DOMAIN only if it has been published more than a number of years ago; this number of years corresponds to the specific national country workflow **pub_term** parameter.

b) In case the AV work has not been published (*No arrow, which corresponds also to N/A input value*) the copyright status algorithm deduces that the AV work is in the PUBLIC DOMAIN.

c) In case a premiere date is not found (*Unknown arrow*) and it cannot be established if the AV work has ever been published the algorithm cannot determine the copyright status and the UNDEFINED value is returned.

This is the generic algorithm. Obviously changes or customisation to the algorithm can be applied later in the project based on the requirements of the partners. For example, one customisation would be that of considering some UNDEFINED results described previously as ASSUME COPYRIGHT or PUBLIC DOMAIN.

The next figure depicts the sub-algorithm that establishes the copyright status of a AV work that is not anonymous, based on information related with its authors. As defined in the Glossary, the work authors are the ones that contribute to the creation of a cinematographic or AV work in any of the following roles:

- principal director
- author of the screenplay
- author of the dialogue
- composer of music specifically created for use in the work

The roles of the contributors that are considered work authors are also referred to in this document as "Main Contributor Roles". Although, at the current state, the list of authors considered "Main Contributor Roles" is the same throughout all EU countries (harmonised by EU directive regarding of copyright protection term), keeping such list configurable provides better flexibility for the FORWARD workflow in case of future changes. It's worth pointing out also that not all the roles maybe applicable to every AV work (e.g. a silent film does not have a composer).

For the algorithm described below, there are the following main input parameters:

- the list of main contributor roles applicable to the work :**Main Contributor roles**
- the list of the work contributors in a main role: **Main Contributor list**
- information whether there are any main contributor roles for which a contributor could not be found : **missingContributorsWithMainRole**
- an auxiliary flag which registers for every know contributor whether his/her death date is found or not **unknownContributorExpiry**
- national country workflow and its related rules in calculating expiry term for authors : **National Workflow box**

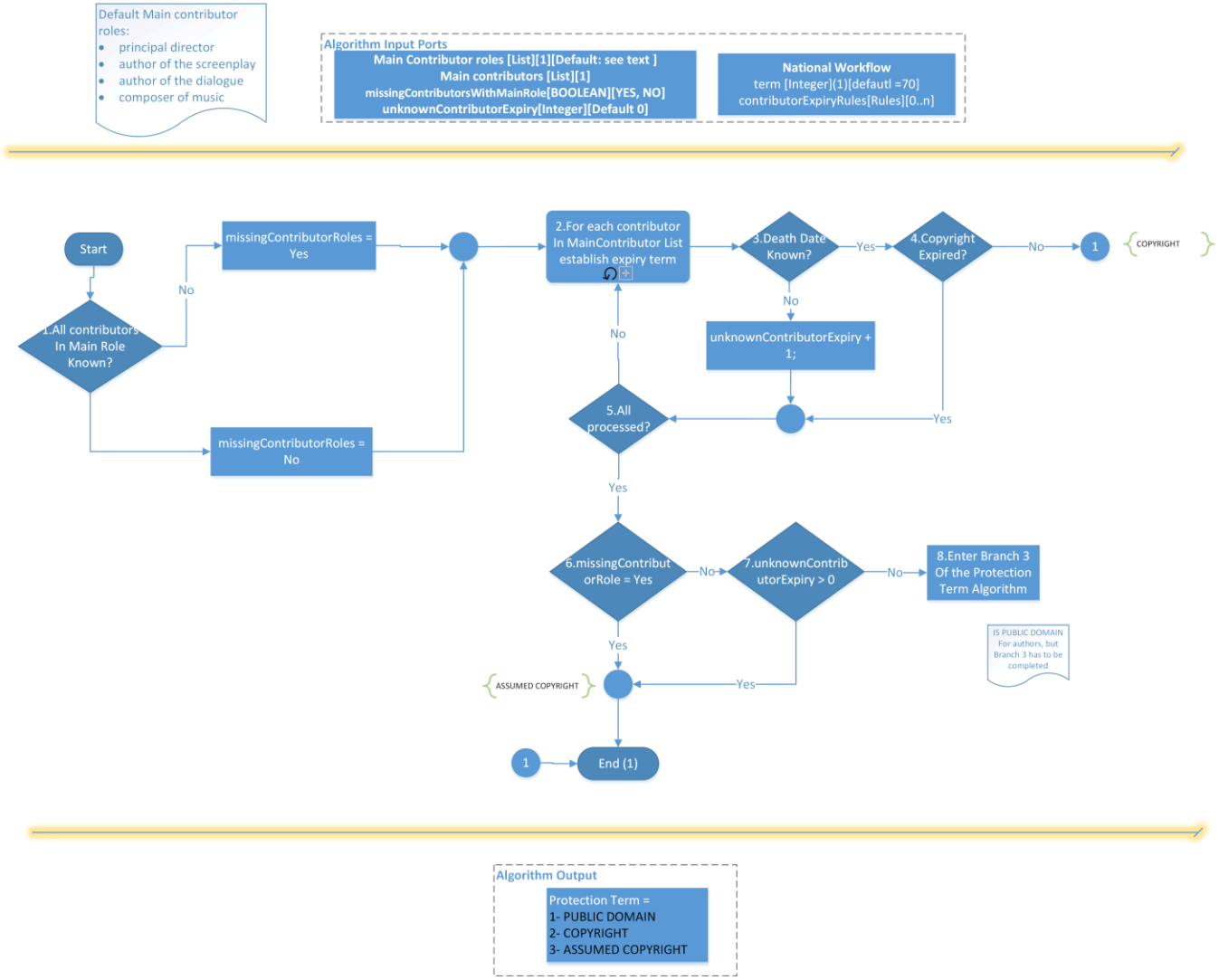


Figure 4 – Copyright Status Algorithm (2/2)

The first step of the above algorithm checks whether there is any contributor having a main role that could not be found. Even if this information is not used immediately it is important later in the decision diamond having label 6.

Although national variations may exist in determining the expiry copyright date for an author (different term periods, or different terms period based on the author properties such as authors having who died in war), all of them require the death date at some point of the calculation for copyright expiration. This is why for each author, the first check performed concerns her/his date of death.

- If the death date is known, the second check concerns finding out whether copyright has expired for the corresponding author (Step 4). The most common case in determining copyright expiry for an author is that of checking whether the **author death date** + **term** > **year now**, but national variation can exists.
    - If the copyright has not expired for **just one author** (No connector from the 4th diamond – Copyright Expired), the work is in COPYRIGHT and the algorithm ends.
    - If the copyright has expired for the author currently being processed, the Loop at step 2 goes on with the processing of the other authors
- In case the death date is not known (Yes connector from the 3rd diamond – Death date known) then the **unknownContributorExpiry** variable is increased by one and the looping among authors goes on.

After all the main known contributors have been processed, we know that:

- the work is not in COPYRIGHT (otherwise the loop would have broken earlier)
- if there was any known death date for the processed contributors, then their copyright has expired.

Nonetheless, it is still impossible to state that the copyright status based on Contributors is PUBLIC DOMAIN. The Public Domain status is obtained only when the protection term has expired for ALL the authors, meaning that all of them are known and the expiry term can be established for all of them. That's the reason why step 6 first checks for any missing contributors.

- In case there are missing contributors it cannot be stated that the work is in the PUBLIC DOMAIN and that's why the ASSUME COPYRIGHT decision is taken.
- In case there are no missing contributors, but there are some contributors for whom the expiry (death) date is not known, it cannot be stated that the work is in the PUBLIC DOMAIN and the ASSUME COPYRIGHT decision is taken.

In case all main contributors are known and expiry of copyright can be established for all of them, the work is in the PUBLIC DOMAIN with regard to the copyright protection term based on authors. Afterwards, the control is returned to the parent algorithm (Fig.3) which should proceed with Branch 3 evaluation.

1.2.2 Copyright Status Workflow

The above algorithm provides the first base for modelling the FORWARD system workflow. In order to fully allow assessment of the rights status of an AV work the FORWARD workflow takes into consideration also the following main aspects:

- National variations in terms of law and operating procedures
- Missing information in the metadata describing AV works for which a diligent search is needed.
- Available Resource Analysis
- Diligent search tracking
- Flexibility and Extensibility of the workflow

Each of these aspects will be highlighted while describing the FORWARD Workflow for Copyright Status in the current section and subsections. The first figure shows the FORWARD top level workflow for the copyright status assessment. For the sake of clarity some of the processes and interactions are detailed in separate figures.

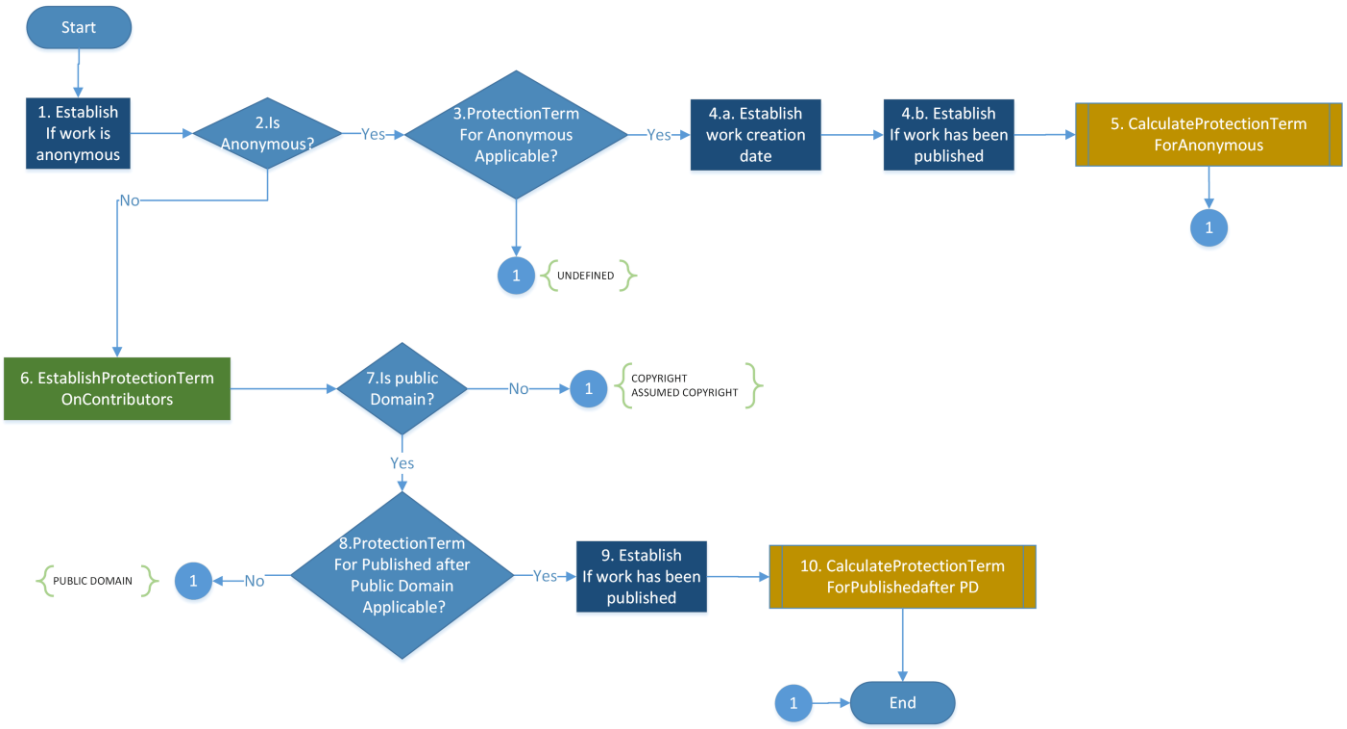


Figure 5 - Copyright Status Workflow

At first glance this workflow is similar to the algorithm described in Figure 3, but there are of course some added elements that stand out such as:

a) The appearances of dark blue boxes that start with the "Establish" verb (step 1, 4, 9)
b) The appearance of some decision diamonds (step 3, 8)

Let's look at more in detail as to what they refer.

In section 1.2.1, as the first step, the algorithm asked if a work is anonymous or not. In order to provide the answer to this question a process is required. The complexity of such a process can vary from fairly simple (e.g. in case the information is present in the input provided by the user) to fairly complex in case an external diligent search source is required. **Diligent Search** is the process required to establish the value of an element necessary for the workflow to go on. Such a process can refer either to searches performed internally in the FORWARD system (provided metadata or FORWARD repository) or to searches performed on external resources. By convention, in all the following workflow representations, all the dark blue boxes used in the diagrams from (in the current figure steps 1, 4, 9) refer to the diligent search processes required to establish the value of the element specified in such boxes. The diligent search workflow is explained in detail in **Fig.14 and Fig.15**, but at the present stage it is important to highlight that:

- the FORWARD workflow includes processes to establish the value of elements that are necessary for determining the rights status of the work.
- the fact that the Diligent Search complexity can vary is also an evidence of the workflow flexibility to handle different national scenarios (e.g. there can be countries that provide the anonymity value of an AV work in the metadata and there maybe countries that really need to perform an external diligent search. The same country may need both scenarios based on the information in possession of the

beneficiary).

- the FORWARD workflow establishes a parametric diligent search engine that can be exploited to serve different diligent search needs as further explained in **Fig.14 and Fig.15**. This parametric engine thus provides the FORWARD system workflow with extensibility and reusability in case diligent searches are necessary for other additional elements of the workflow.
- Workflow component reusability: the same diligent search process can be attached to different workflow steps. See for example that Step 4 and 9 refer to the same diligent process related with establishing of the **premiere_date** of an AV work.

The above FORWARD workflow diagram also introduces the decision diamonds labelled 3 and 8. Introduction of such steps provides workflow flexibility as well, since it allows any national workflow to decide if a particular workflow branch has to be carried out or not. For example, considering the countries analysed so far, the Yes branch of the 8th decision diamond does not always have to be applied.

The other components depicted in the above diagram refer to:

a) the FORWARD Workflow to establish the copyright status of works with known contributors – Step 6. It is elaborated in the next subsection.
b) FORWARD procedures that calculate respectively the copyright status of anonymous works (Step 5) and of works published only after falling in the public domain (Step10). They are described in detail in sections 1.2.2.2 and 1.2.2.3.

1.2.2.1 Establish Copyright Status for works with known contributors

This section describes in more detail the FORWARD Workflow in charge of determining the copyright status based on Main contributors/authors. Step 12 and 16 are described separately in Fig. 7 and Fig. 8 respectively.

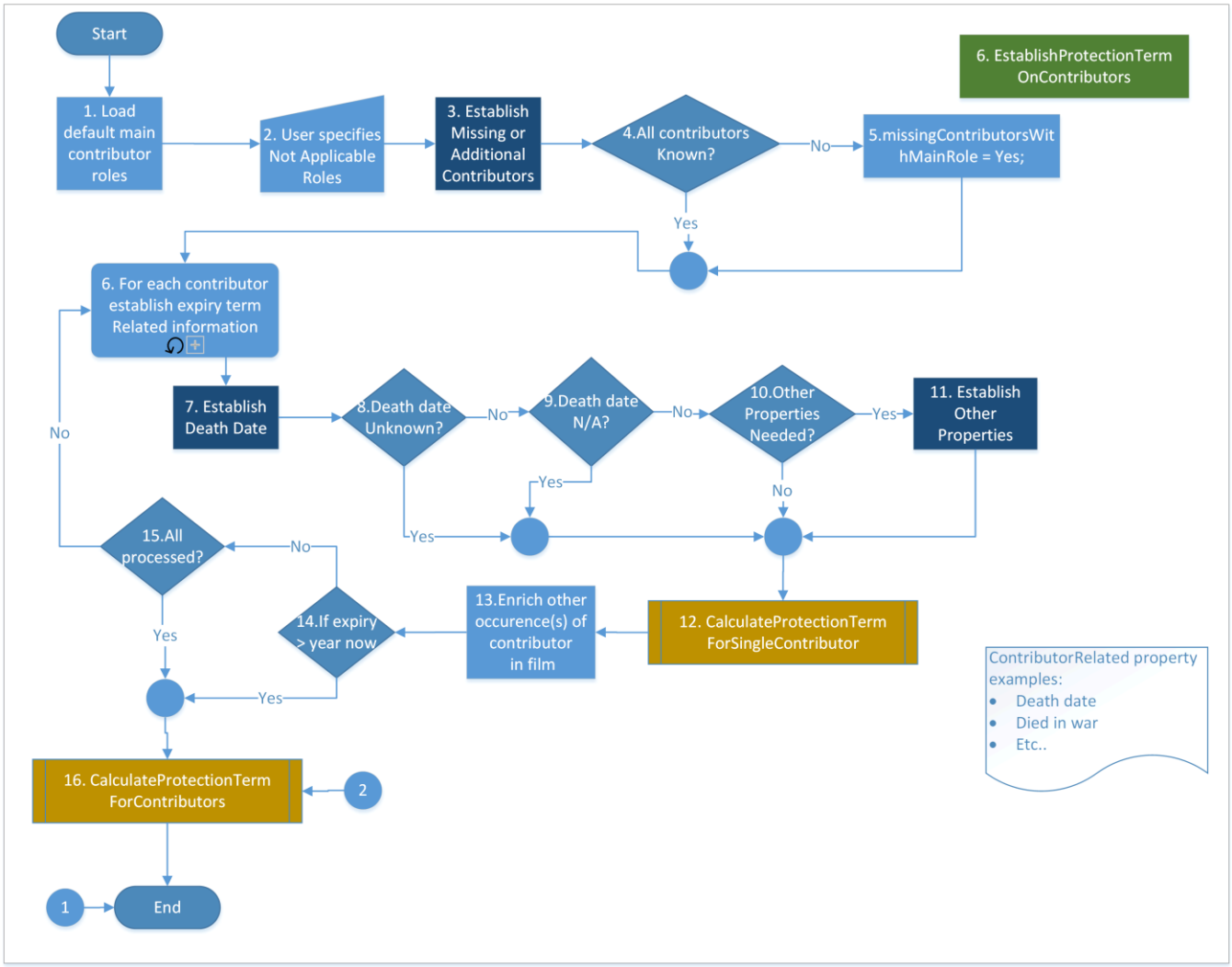


Figure 6 - Copyright Status Workflow

The first step in establishing the copyright status based on authors is that of determining the contributor roles that are considered authors ("Main contributors") when calculating protection term expiration. As mentioned above, such roles usually include principal director, author of screenplay, author of dialogue, composer of music. For each country workflow the FORWARD system will have a default contributor list. While processing the AV work metadata the system will identify if the incoming request does not specify the contributor in any of the main roles. If there is any main role with name of contributor missing the system will ask the user to specify the reason why such main role contributor is missing; in practice the user will specify one of the following choices:

- The contributor with main role is missing since it is not applicable (e.g. the composer in a silent film), or
- The contributor should be applicable and the information is missing.

In the latter case, a **Diligent Search** is performed in order to establish both main role missing contributors and additional authors in the same role (e.g. two directors). If some author is not found as a result of the Diligent

Search, the workflow request variable **missingContributorsWithMainRole** is set to Yes.

The workflow continues at Step 6 where each of the found contributors will be analysed in order to establish the information that is necessary in determining their copyright expiry. The first information to be retrieved is obviously the death date (Step 7).

- In case the death date is unknown (Step 8) or Not Applicable (Step 9 – author is still alive) the flow calls the **CalculateProtectionTermForSingleContributor** procedure (Step 12).
- If date of death is found, before calling the **CalculateProtectionTermForSingleContributor** procedure, based on the workflow country instance, the system evaluates if there is any other contributor property configured that should be taken into consideration while determining the expiry term for the contributor (Step.10).
  This branch is obviously not requested in all the workflows but it may be necessary for example in the French case when the expiry term for contributors who died for France is different from the ones who died in different circumstances. If such a property is needed a Diligent Search has to be performed.
  In the most generic case one or more properties may be required and separate diligent searches may be performed for all of them. For the sake of not adding additional complexity in the above figure just one additional property besides the death date is represented, but more than one can be configured if necessary.

The elaboration performed at Step 12 is explained in Fig. 7 but for the scope of the description of the workflow represented in this figure, the important information is the output provided by the **CalculateProtectionTermForSingleContributor.** The output is one of the following: *Copyright Expiry Date* for that contributor or *Unknown* when such information could not be established.

In case the system detects that the currently processed contributor appears more than in one role in the work being processed, its information can be updated so the user does not have insert such information again.

At step 14 the system checks if copyright of the contributor being processed has already expired or not by comparing it with the current year.

a) If copyright has not expired, the loop at Step 6 is interrupted and the procedure **CalculateProtectionTermForContributors** (explained in Fig. 8) is immediately called. This shortcut can be taken because identifying at least one main contributor, whose copyright has not yet expired, is sufficient for declaring the AV work to be in COPYRIGHT. This way, the FORWARD system and its user do not have to carry on uselessly the same steps for all the remaining contributors. The process could be further optimised by first checking the copyright of those contributors for which there already is a death date available.. During the design and implementation phase further optimisations can come to light.
b) If copyright has not expired the loop at Step 6 continues with processing the next contributor.

Step 16 – responsible for calculating the AV work Copyright Status based on contributors is called whenever the loop at Step 6 is interrupted as in case a) above, or when the loop ends when all the main contributors have been processed.

Next, the procedures that calculate the Copyright Status for a Single Contributor and the AV work Copyright Status based on the information gathered from the workflow will be described.

Calculate Protection Term for Single Contributor

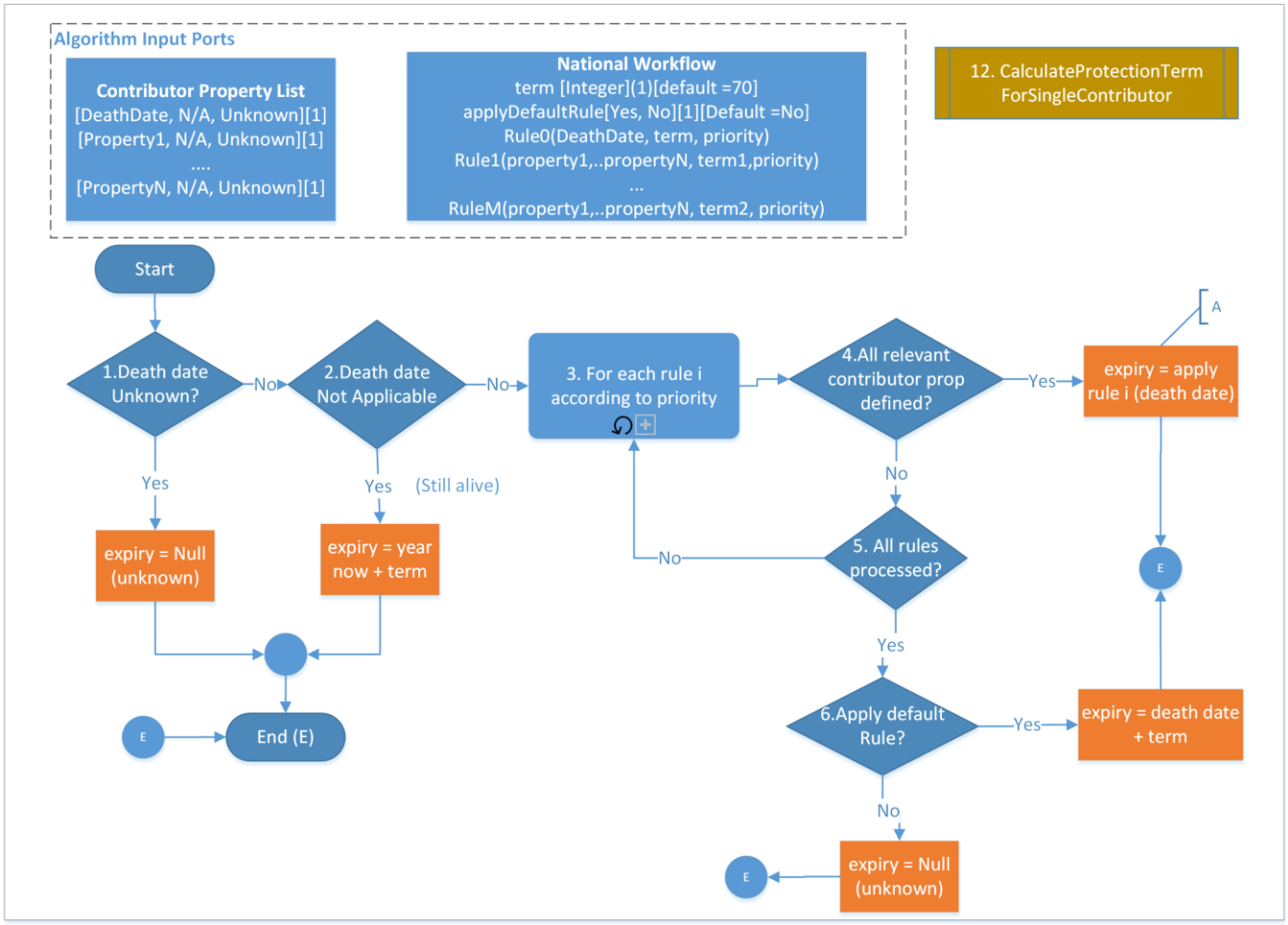


Figure 7 - Determine max expiry year for single contributor procedure

Let's explain briefly the **Algorithm Inputs**.

The **contributor property list** contains all the contributor-related information requested by the current workflow instance in order to establish the copyright expiry. They obviously include the death date (that is part of these properties, but is explicitly stated since it is the core property) and other properties (such as died in war etc.).

Another property not used for now is the contributor birth date. Where the death date may be unknown a contributor's birth date maybe taken into consideration for establishing the copyright expiry date. Obviously in this case the copyright term applied will be different.

For each property one of three values is expected:

- Its value when known (the death date, died in war (yes, no) etc.)
- Not applicable when it is not applied (for example death date in case of a still living contributor)
- Unknown – when it was impossible to establish it.

The **national workflow parameters** contain:

- the **term** used by the workflow which indicates the number of years after which the copyright for a main contributor expires.
- A **list of rules** to be applied when calculating copyright expiry. Each rule has:
  - A list of contributor properties used in the calculation
  - The **term** to be applied in that case
  - A priority that establishes the order in which the rules will be evaluated.
- **applyDefaultRule**, a workflow configuration parameter used to evaluate if the default rule (death date + term) should be applied in case the information that is required for the other rules is not available.

Once the procedure obtains the **Contributor Property List,** it checks for an unknown death date. In the affirmative case the expiry for that contributor is set to Unknown. In case the contributor is still alive, the Not Applicable (N/A) is provided, and the expiry is set to a date that is in the future. Anyway, this operation should not to be taken literally; the main point is that in some way the Contributor Expiry Algorithm will distinguish between Unknown and Not Applicable.

In case a death date is present, the system evaluates all the contributor expiry rules according to their pre-configured priority.

- Whenever the necessary information requested by a rule is present, the rule term is applied (orange box on the top rights labelled A), otherwise the next remaining rule with the highest priority is evaluated. (No-connector after Step 5).
- If none of the rules can be applied (Yes-connector after Step 5), the national country workflow configuration (**applyDefaultRule**) is checked to see if applying the default rule is allowed, and expiry is calculated as: expiry = death date + term. If the default rule should not be applied, unknown is provided. For example, the French workflow may be configured to apply the default rule whenever the contributor death date is known but the information whether the contributor has died for France is not known.

The introduction of the property list and the national workflow rules that can be used is a generalisation so that it allows the algorithm and the FORWARD workflow to be easily extended and configured in case national needs require it.

The result of the algorithm the expiry date will be UNKNOWN or a Date.

Calculate Copyright Status based on contributors

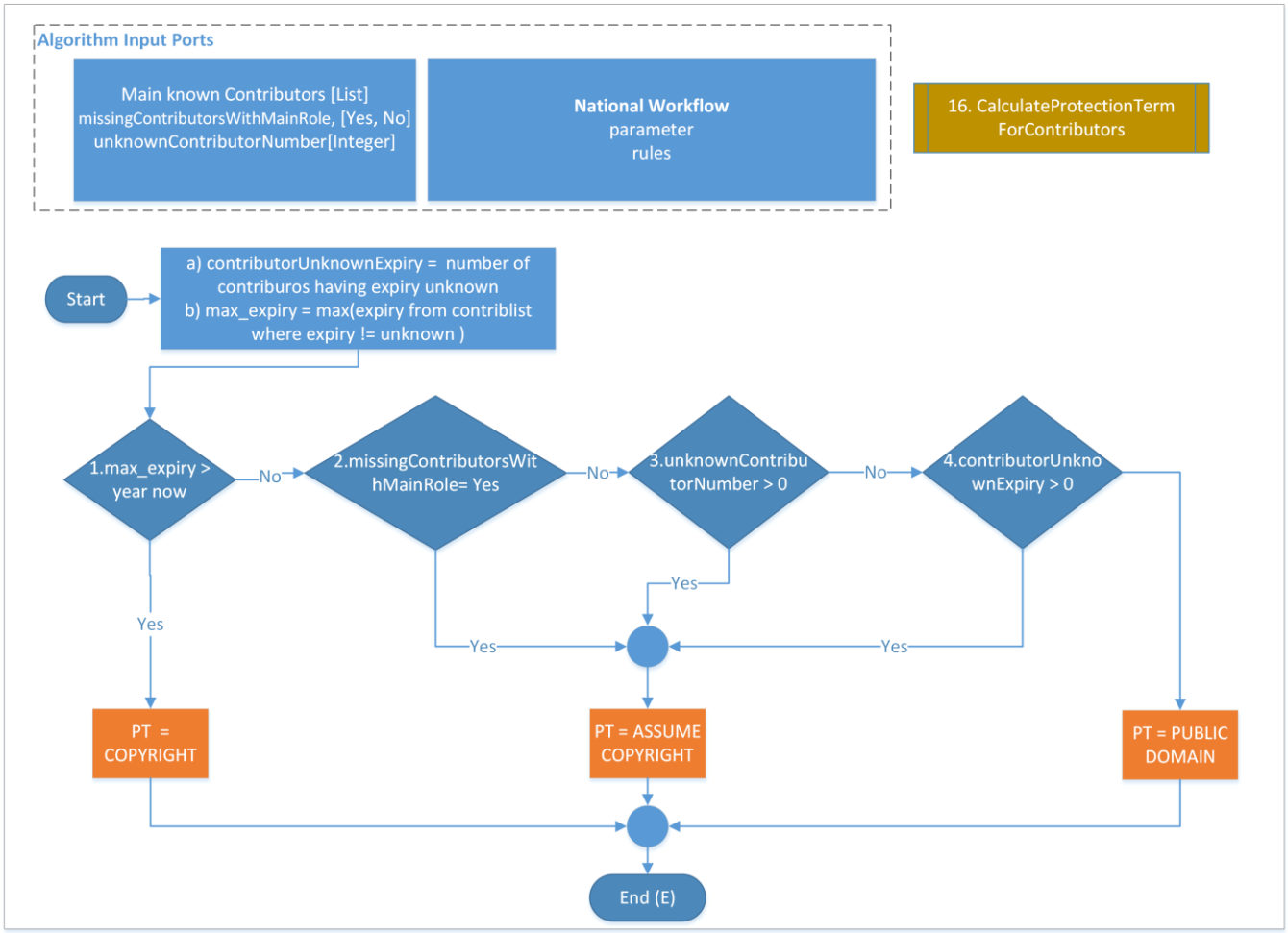


Figure 8 - Determine Copyright Status based on Contributors Procedure

This procedure will not be described in detail since it is quite similar to the copyright status algorithm based on contributors described in 1.2.1. The result of the algorithm – the copyright status – will be COPYRIGHT, ASSUME COPYRIGHT or PUBLIC DOMAIN.

1.2.2.2 Calculate Copyright Status for Anonymous Works

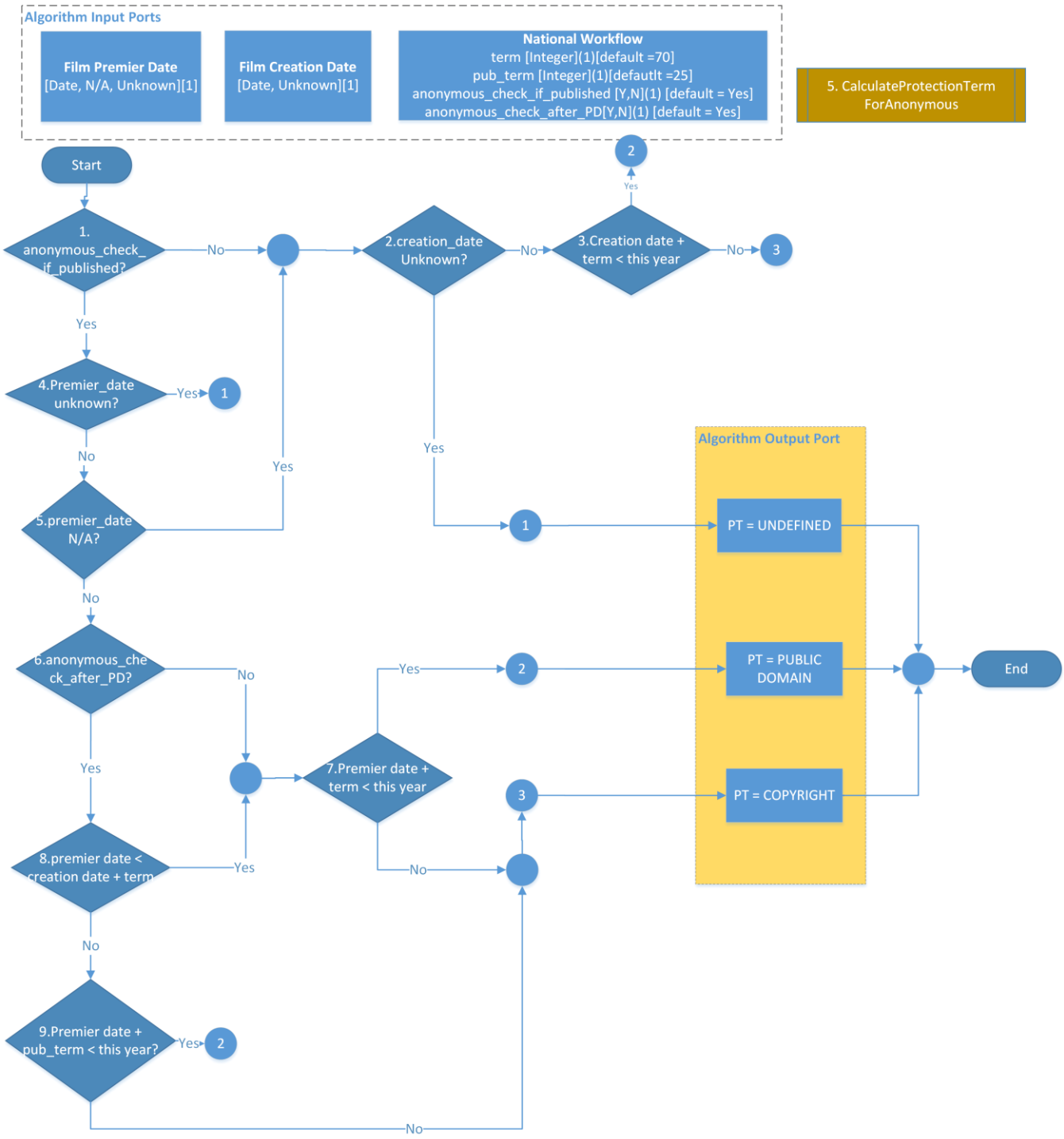


Figure 9 - Copyright Status Procedure for Anonymous Works

The above procedure corresponds to the Step 5 indicated in Fig. 5, and describes the way the copyright status

for anonymous works is calculated. The logic here follows the one explained in branch 1 of the Copyright Status algorithm in 1.2.1. The only difference that can be noticed here is the introduction of two workflow parameters that provide flexibility to the FORWARD workflow. They address the existing national variations in the management of the protection terms for anonymous works. In particular they are:

- **anonymous_check_if_published** – indicates whether that country workflow needs to consider the publication date in establishing the protection term for anonymous work.
- **anonymous_check_after_PD** – indicates whether that country workflow needs to consider whether an anonymous work was published after having fallen into the public domain or not.

1.2.2.3 Calculate Copyright Status for Works Published After PD

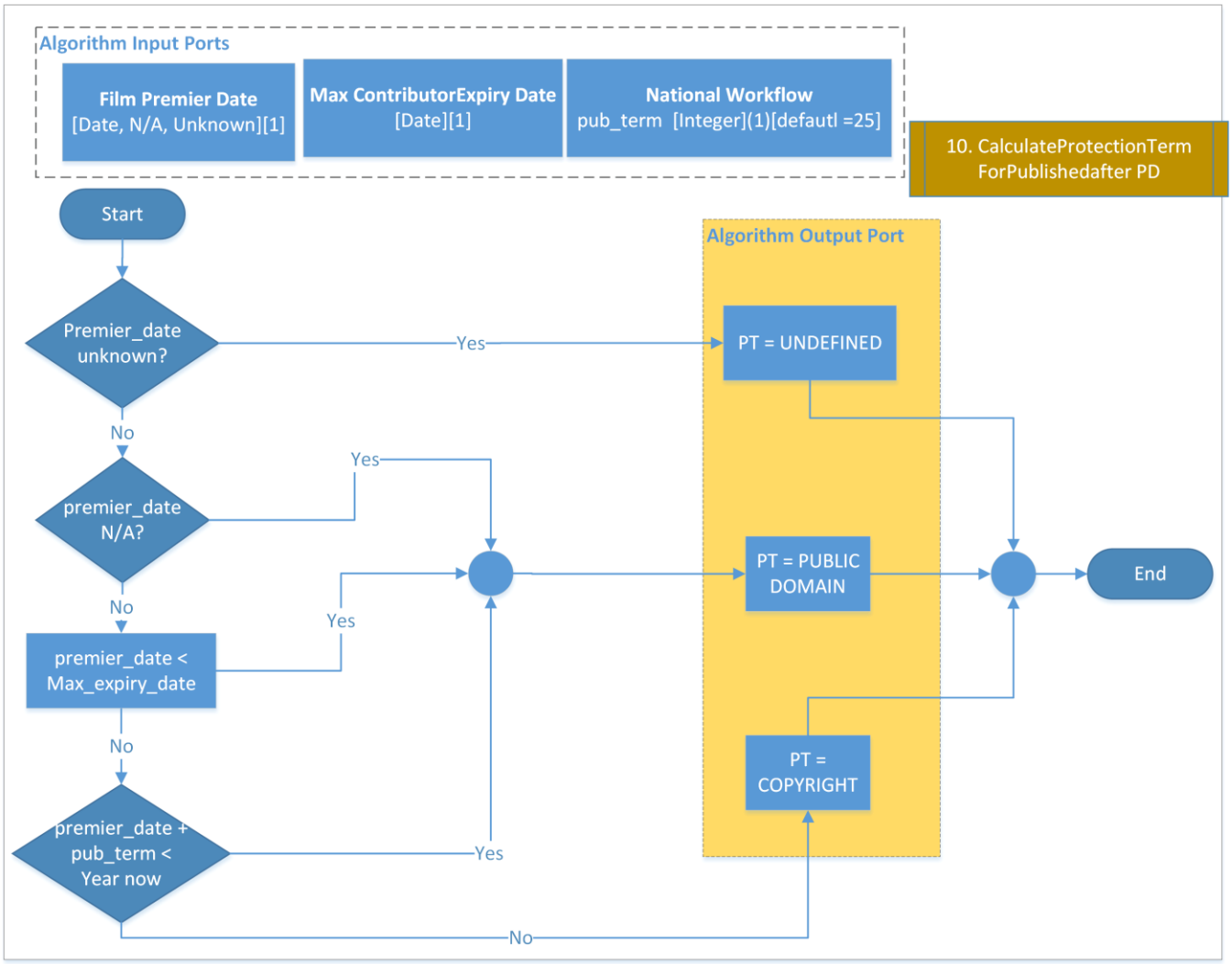


Figure 10 – Copyright Status Procedure for Works Published After Public Domain

The above procedure corresponds to the Step 10, indicated in Fig. 5 and describes the way the Copyright Status is calculated in this case. The logic here is the same as in branch 2 of the Copyright Status algorithm in 1.2.1.

In case the **premiere_date** is unknown the Copyright Status of the AV work is considered UNDEFINED. As stated earlier in the document, considering that this branch is applied only if the main contributors' copyright expired,

based on the user's decision, the copyright status in this case could also be considered PUBLIC DOMAIN.

If the **premiere_date** is NOT applicable, meaning the AV work has never been published, the work is in the PUBLIC DOMAIN.

If the **premiere_date** < **max_expiry_date**, meaning that the work has been published before the copyright on main contributors expired, the work is in the PUBLIC DOMAIN. In the other case, the work has been published after the copyright of main contributors has expired and it is declared to be in the PUBLIC DOMAIN only if it hasn't been published in the last **pub_term** years.

Please note that the max contributor expiration variable will always be present when this procedure is run, since the max contributor expiration is established at the end of workflow Step 6 (Fig. 5)

1.3 Rightholders Determination – Establish Orphan Status

As in the previous chapter, before describing the FORWARD System Rightholders Determination – Establish Orphan Status workflow, the logical decision tree/algorithm needs to be described.

Please note that the logical decision tree describes the decision logic taken based on the provided or gathered information, it does not describe the way this information is retrieved. The way the information is retrieved is described in the related workflow.

1.3.1 Rightholders Determination – Establish Orphan Status Algorithm

A work is established to be ORPHAN when none of its rightholders can be identified and located after a diligent search. As depicted in the figure below, there are two main necessary phases for establishing whether the work is ORPHAN or NOT.

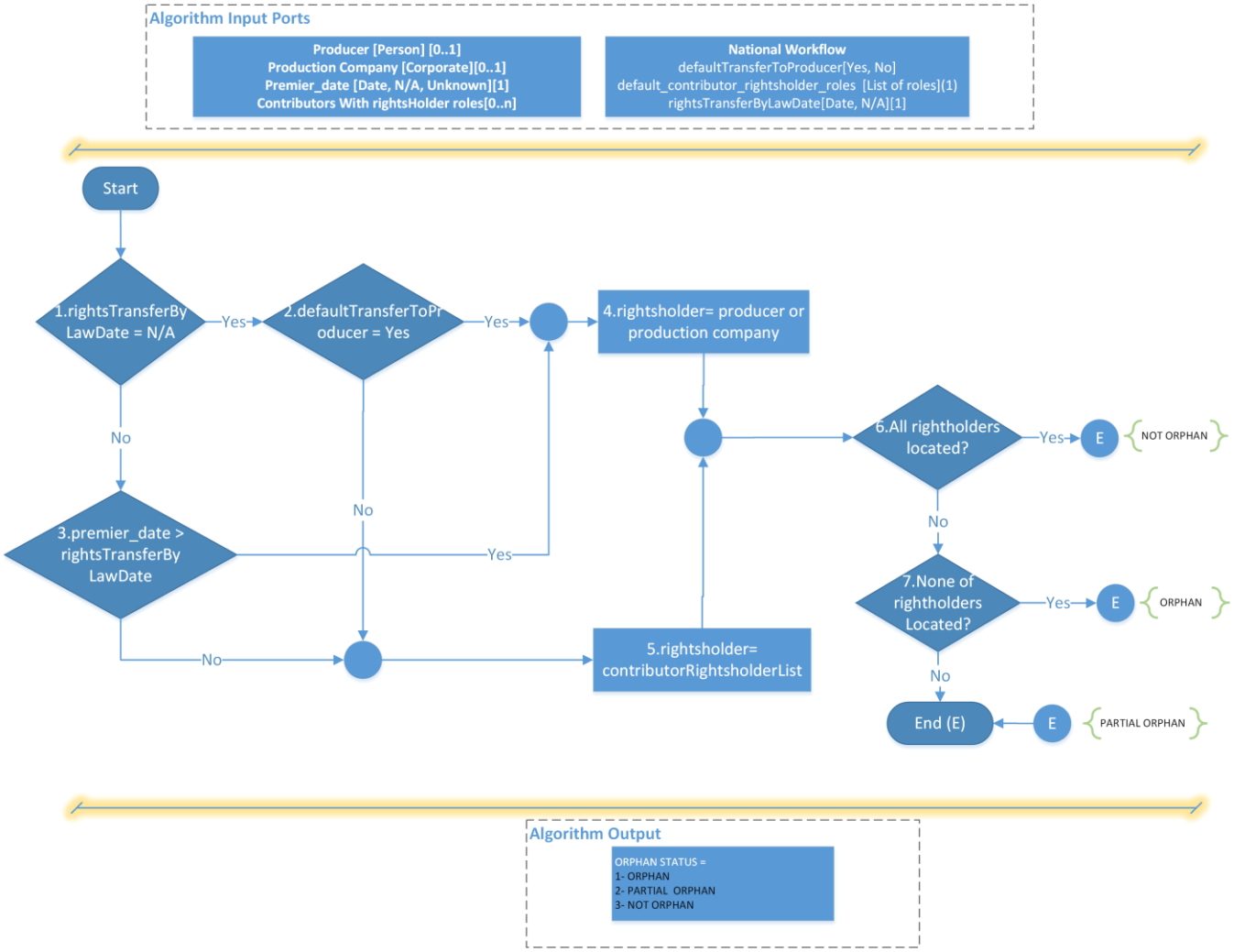


Figure 11 – Rightholders Determination – Orphan Status Algorithm

Phase 1: Determine Rightholders:

The aim of this phase (covered by steps 1–5 in the above figure) is that of establishing who are the rightsholders of the work. Basically, they can be the producer(s)/production company(ies) or a list of contributors who have a rightsholder role. This determination varies from country to country: e.g. in some countries the rights are transferred by law to the producer/production company if the AV work was published after a certain date, while in other countries the rights are transferred to the producer/production company by default. This is why the first step of the algorithm checks if a rightsTransferByLawDate is applicable or not for the national workflow.

- In case such law is applicable (No connector 1st diamond), the **rightsTransferByLawDate** is compared with the AV work **premiere_date**. In case the AV work has been premiered after such a date the rightholders are determined to be the producer or the production company. Otherwise the rightholders are determined to be the contributors with rightholder roles.

  The **rightTransferByLawDate** and the role list of contributors who have rightholders role are also subject to the national workflow configurability.

- In case such law is not applicable (N/A), a check whether there is a default transfer of the rights to the producer/production company is performed. In the affirmative case the rightholders are determined as the producer/production company. Otherwise the rightholders are determined as the contributors with rightholder roles.

Phase 2: Locate Rightholders

Once the set of rightholders has been established, the determination whether the work is ORPHAN or not, is based on the number of the located rightholders.

- If all the rightholders are located, the work is NOT ORPHAN
- If just a subset of rightholders is located, the work is PARTIAL ORPHAN
- If none of the rightholders is located, the work is ORPHAN

1.3.2 Rightholders Determination – Establish Orphan Status Workflow

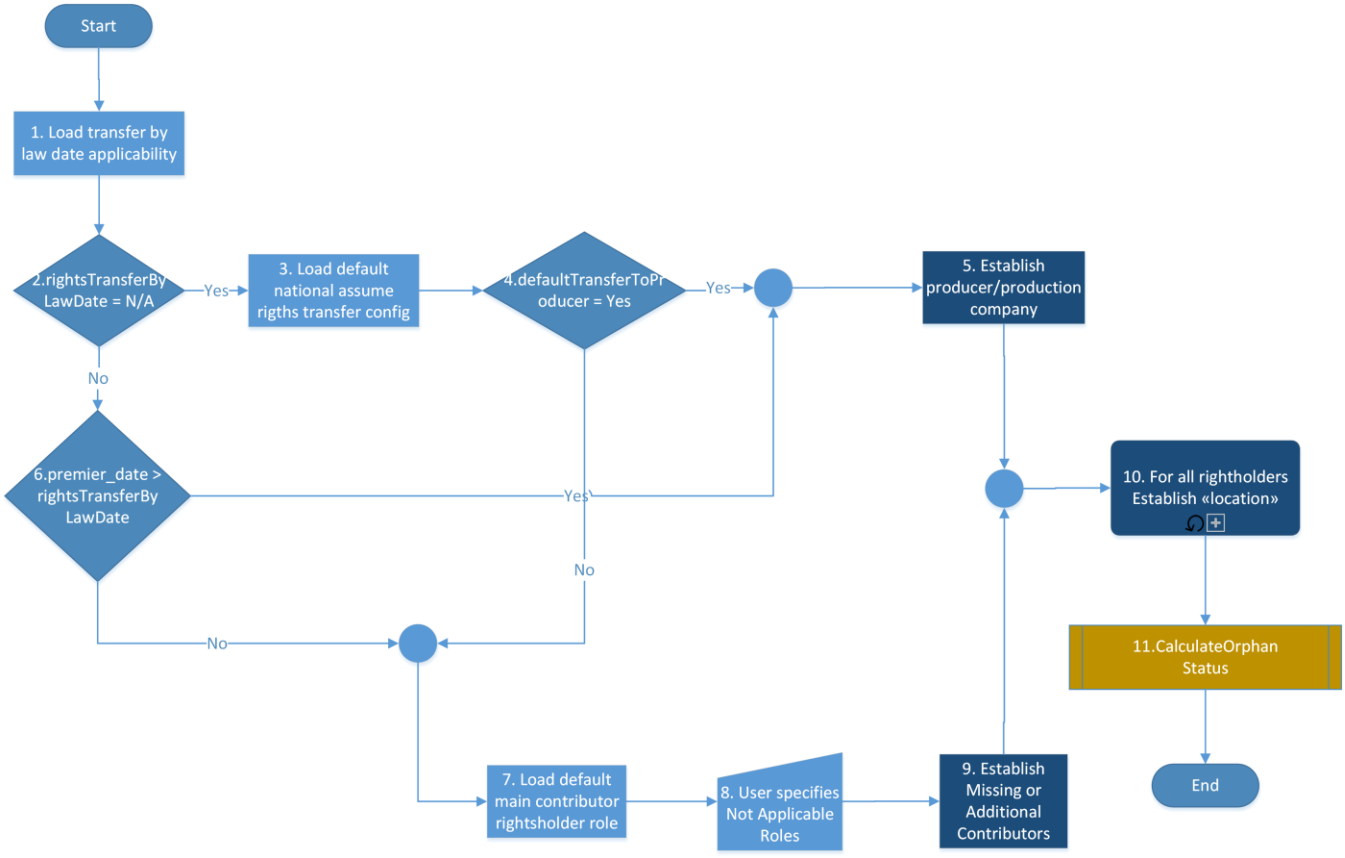


Figure 12 – Rightholders Determination –Establish Orphan Work Workflow

At the beginning of the workflow the system checks if the national workflow contains a transfer by date law. In case it does, the workflow checks whether the AV work has been published after the rights transfer by law date to the producer (Step 6). In this case the AV work rightsholder is the producer or the production company,

otherwise the system loads the default contributor roles which are eligible for being rightsholders. This list is subject to the national workflow and the AV work itself. For this reason, after loading the default contributor list with rightholder role, the user has to establish if any of those roles are not applicable for that particular Work.

Once the rightsholder role is established, a diligent search is performed in order to establish if the rightsholders are known (Step 5 or Step 9 according to the case). “Known” in this case means having *at least* the name of such rightsholders.

Next, a diligent search is performed to locate the known rightsholders (Step. 10). In case the rightsholders is dead or does not exist anymore (in case of a corporate rightholder), the act of locating rightsholders may require searching for heirs/successors, as prescribed by national law.

When the diligent search for their location is finished, the Calculate Orphan status procedure is called. The procedure is described in the next subsection.

Please note that in Step 6 it is expected that the premiere_date has been established at this stage of the process since it is required in previous workflow steps. However, in case the premiere_date is not know, a diligent search on it can be added to the above workflow.

1.3.2.1 Calculate Orphan Status Procedure

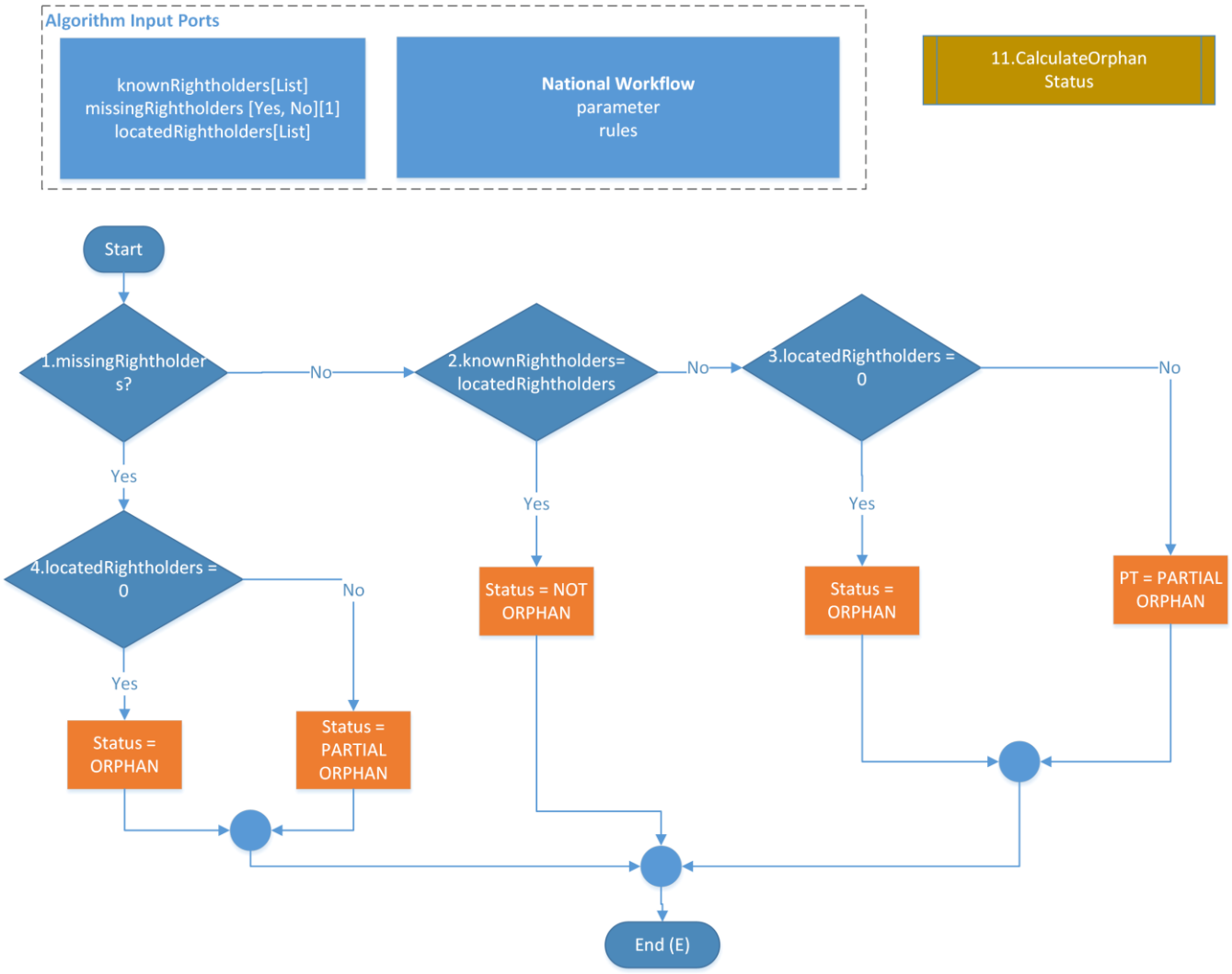


Figure 13 - Calculate Orphan Status Procedure

The above diagram describes the procedure for calculating the Orphan Status.

Initially the procedure checks if all the rightholders are known:

- In the affirmative case, if all the known rightholders have been located, the work is NOT ORPHAN. If none of the known rightholders have been located (Step 3), the work's status is ORPHAN, otherwise it is PARTIAL ORPHAN.
- If not all rightholders are known (Step 4), the work is ORPHAN if none of the known rightholders have been located, otherwise the work is NOT ORPHAN.

1.4 Diligent Search Workflow

This section describes the generic Diligent Search Workflow (DS workflow) that is used in the workflow diagrams of both the Copyright Status and Rightsholders Determination – Establish Orphan Status macro-steps. As stated earlier in this document they refer to each of the dark blue boxes. Based on the input parameters (element

types, corresponding list of source to be consulted, etc.) and obtained results not all of its branches may be performed. It is a generic diligent search workflow engine that should handle all the different situations.

In order to explain the DS Workflow, the swim lane notation is used to show the interaction among system actors since it is important to understand the role that each of them has. From Deliverable 3.1 Sector Analysis and User Requirements it came out that most of the external resources to be queried do not provide an API or a cost effective integration. That explains why the diligent search workflow foresees human intervention in many steps.

The main pattern in the DS workflow is the following:

- If an element is not present in the metadata, perform an internal search in the FORWARD repository when applicable.
- If the internal search yields one or more results, the user has to analyse them and either select one or reject all of them.
- In case no result is found as a result of the internal search, an external diligent search is performed.
- Even in this case, if results are found they will be presented to the user for validation and selection.
- In all cases the system stores the decisions taken. For each decision the following information shall be saved:
  Request Id; Workflow Id, user, selected information, source(s) where it was found, date and notes.
  This is a preliminary list but it provides an initial idea of what diligent search involves.

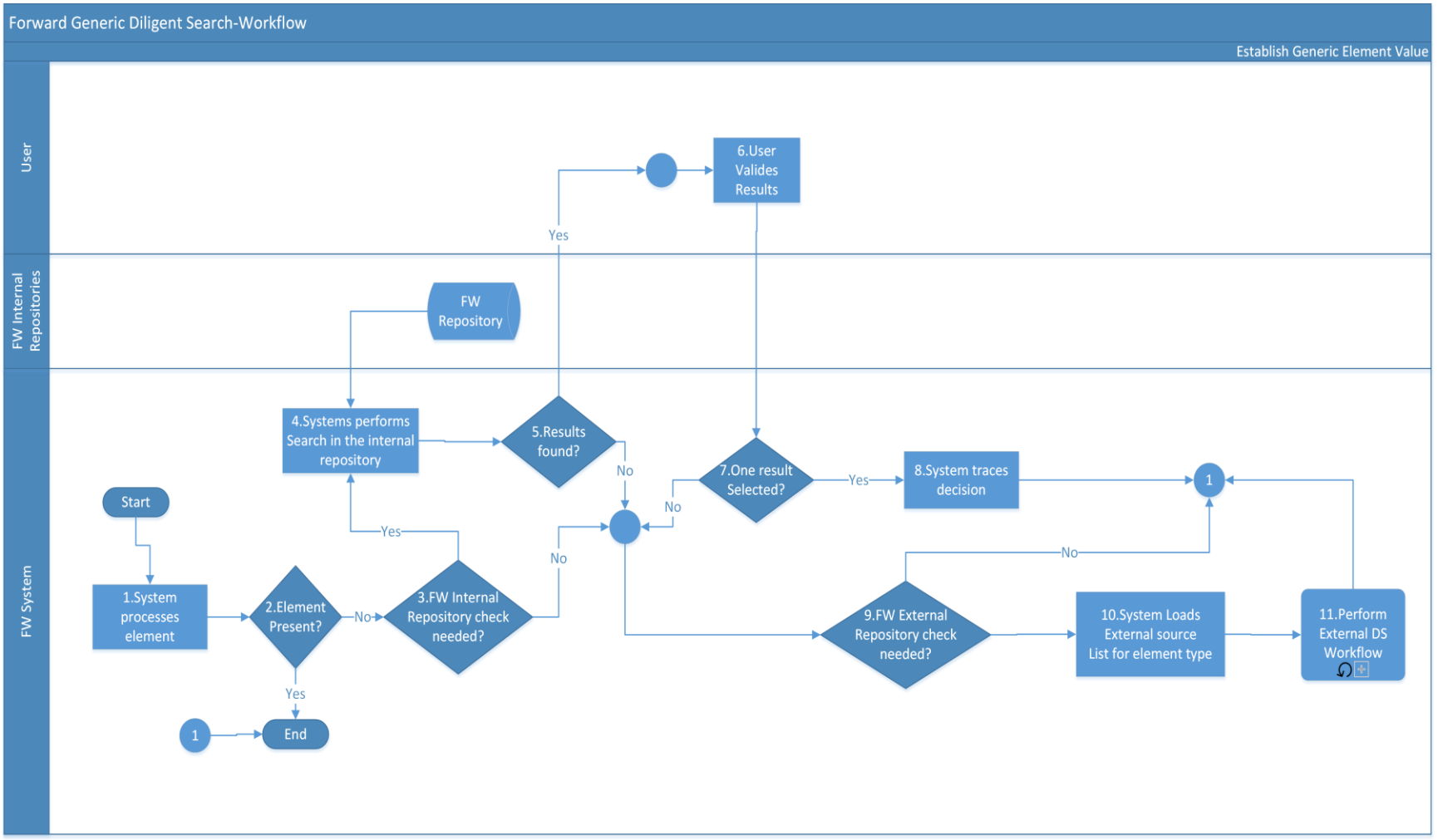


Figure 14 - Generic Diligent Search Workflow

The figure above describes the generic diligent search workflow. It focuses on the interaction between the FORWARD System, its users and the Internal FORWARD repository. The Diligent search performed on external data sources (Step 11) is described in the next diagram.

Whenever the FORWARD workflow needs an element, the first step of the DS workflow is to check if the element is present in the metadata provided by the user (Step.1).

- In the simplest case, the information is present and the diligent search workflow ends.
- In case the information is not present, based on the element and the national workflow, the system checks (Step 3) whether an internal search in the FORWARD repository is necessary before querying external sources. For example, in case the system does not have the contributor death date, before performing the external diligent search the system might check if the requested information for the contributor in question is already present in the FORWARD repository due to a previous process. In case the premiere date of an AV work is missing, the internal search may not be necessary and only the external diligent search has to be carried out.
- In case an internal search is necessary the system searches the necessary information in the FORWARD repository. In case one or more results are found, the system asks the user to validate them. If the user selects one of the proposed results the system stores the decision and the diligent search workflow ends. The selected value is returned to the workflow step that invoked the DS workflow.
- In case the internal diligent search is not necessary, or it is necessary but the user found no satisfactory results, the external diligent search is performed where applicable.

In case of an external diligent search, the FORWARD workflow loads the external source list to be queried based on:

- **type** of information been requested (person or corporate contributor, person or corporate rightholder, premiere date, person or corporate rightholder etc.)
- information properties (such as contributor place of birth for example)
- national workflow configuration.

The next diagram (Fig. 15) describes the External Diligent Workflow. It should be quite clear and the key aspects are highlighted.

The external diligent search involves two categories of resources:

- Automatic Sources – the ones that can be queried automatically by the FORWARD system
- Manual Sources – the ones that cannot be queried automatically, but should be consulted by the FORWARD user manually.

Some resources may be shared by all the workflows and others are specific.

When an external diligent search is started the system queries in parallel all the automatic sources and displays the aggregated result to the users (Steps 3–11) for validation. If the FORWARD user selects one of the results as the one he was looking for, the decision is stored and the workflow ends.

In all cases, the FORWARD system displays to the user the list of Manual sources that can be consulted together with the automatic search results when found. In case the user wants to consult the manual sources only, he can reject the automatic search results. The Source Link displayed in Step 13 will also contain the links to the automatic sources, in case the automatic search is not giving the expected results and the user wants to perform more specific searches in the systems.

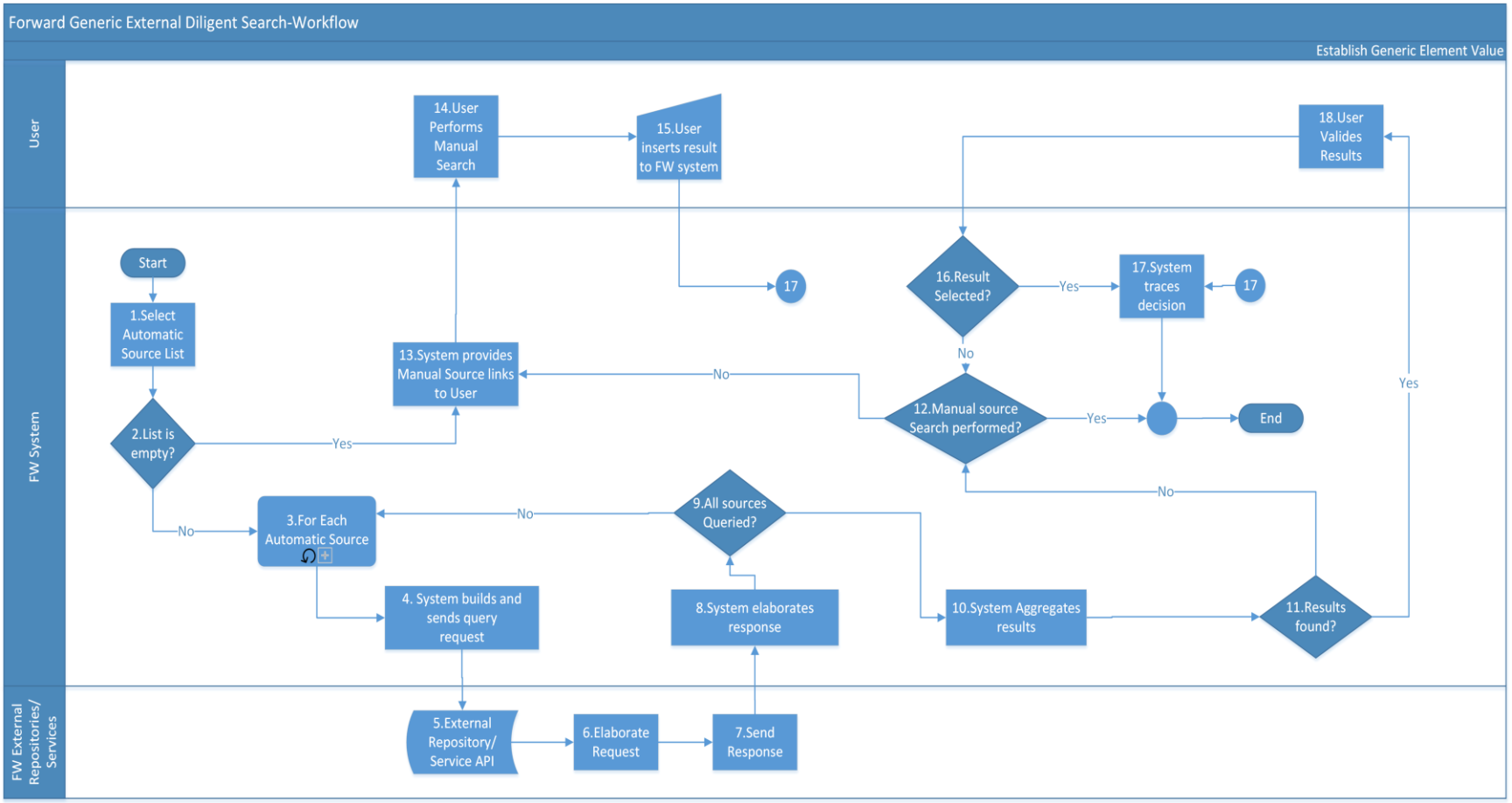


Figure 15 - Generic External Diligent Search Workflow

1.5 FORWARD Workflow Configuration Parameters List

The aim of this subsection is to provide a summary of the workflow configuration parameters requested for each country.

The workflow configuration parameter list has to be filled by each FORWARD country. In order to facilitate easy distribution to the FILM HERITAGE INSTITUTIONS, in charge of filling it, the list is provided as an Excel file. See Annexes FORWARD Workflow Configuration Parameters file. It is important to highlight here that not all the lists contained in the document can be considered as final. They are still under evaluation, or being finalized at national level. As such, the lists are "works in progress" and will be updated along the timeframe of the project.

1.6 FORWARD Workflow vs Arrow Workflow

The differences between the ARROW and FORWARD **workflows** are mainly driven by the following factors:

- Rights Clearance starting unit (work vs manifestation)
- Quality and comprehensiveness of the workflow input data
- Internal and External Data Sources characteristics
- Degree of management of law national variations
- Specific system requirements
- During the ARROW and ARROW projects the Orphan Works Directive was not yet approved.

Let's analyse the differences and highlight their impact on the workflow.

Rights Clearance starting unit (work vs manifestation):

ARROW Users started their diligent search by submitting a Manifestation and one of the main challenges of the system was the clustering process. Clustering in ARROW aimed at identifying the Work against its Expressions and Manifestation and was quite a complex process.

In FORWARD, the whole issue of Work vs Manifestations and Expressions is (almost) completely irrelevant in the AV world because FHI have been using for decades the filmographic information pertaining to the Work (and not to its expressions and manifestations) as the main entry point to their databases. In other words, in FORWARD the query about a given Work will always be initiated at Work level and no clustering is required.

Quality and comprehensiveness of the workflow input data: From the analysis of the legal decision trees and D4.1 "Analysis of Available Resources in the AV sector" mainly highlighted the following:

a) FHIs data do not always have all the information that is necessary to carry out the "*assisted diligent search process*" (ADiSP).
b) There are no widely used standard export metadata schemas in the AV sector whereas in the library world (ARROW) all the libraries used MARC21 and its variations.
c) Even when the information is present, there may be many differences in format and types among FHIs
d) FHIs Authority files in most of the cases do not contain standard identifiers of the entities they represent (contributors for example)

In ARROW the information that was necessary to carry out the Diligent Search process was defined as mandatory and submission files that didn't contain the mandatory metadata were rejected. This choice was due to the ARROW requirement of handling mass digitisation requests where an automatic process was necessary. In FORWARD, the approach of accepting input requests will be less restrictive and the system will provide to the users the possibility to users to complete information during the process. This explains the presence of a lot of diligent search processes in the FORWARD workflow ("Establish element values" boxes).

In ARROW, the Diligent Search started mainly from the libraries which already have their data in MARC21 XML (or other MARC variations) and no particular effort was requested on their behalf on performing data export in such formats. Since there is no equivalent standard widely used by the FHIs in the AV world, in FORWARD it will be necessary to define an export schema for FHIs to export their data in before being processed in FORWARD. FHIs will be responsible for correctly exporting their data in the established schema. Exporting by the FHIs will likely require some data normalization. (Data enrichment – to address issue at point c).

In ARROW, many of the libraries were integrated in VIAF; this means that the records submitted by the library contained the contributor identifiers that significantly facilitated the retrieval of contributor's information (date of death etc.) from VIAF. The use of the identifier not only allowed to easily find the contributor, but it indirectly facilitated the automation of ARROW; by using the identifier just one record was found and there was no need for human intervention to select the "correct" contributor among several. FHIs are not integrated in VIAF, so when the FORWARD workflow has to search for contributor information (even in case of an automatic integration with VIAF was in place) the human intervention in order to choose the rights one is unavoidable.

Internal and External Data Sources characteristics.

In ARROW the main informal sources are external to the system (VIAF, TEL, Books In Print, RROs); they are highly organised and comprehensive, as well as relatively small in number (thanks to a concentration/federation of local databases, like the ones from the publishers).

In FORWARD, the lack of comprehensive, transnational databases imposes the use of a plethora of smaller, national (and even regional, local, personal) databases. The implementation of the Orphan Work Directive in the different member states sometimes include a list of resources to be queried that can be considered more or less mandatory; some of these are useful, some are completely irrelevant. Still, they must all be checked.

Besides the number of the data sources to be integrated, the complexity is further increased by the low interoperability of many of these sources (some of them do not have a B2B interface for example).

Considering the nature of the FORWARD data sources, it is evident that the FORWARD workflow cannot be automatic (as in ARROW), but a lot of human intervention is required. That's why in D.3.2 the FORWARD diligent search is referred to as "assisted diligent search" (ADiSP).

A sum-up of the human intervention ARROW vs FORWARD follows.

ARROW

- In ARROW, the only human interaction was required in the cluster validation phase (in ARROW Plus though the human intervention could be totally avoided under certain configurations and circumstances).

FORWARD

- In many – or most – cases the user has to validate search results retrieved from automatic querying of external sources
- In case no result is obtained from the automatic sources, the user has to manually perform searches in the external sources which do not have an automatic interaction in FORWARD and store the main results obtained in FORWARD.
- The user has to interpret the data (or lack thereof) and take some decision. This does not only refer to a specific search result, but also to the interpretation of a set of results; for example in case the FORWARD system does not have all the necessary elements to rigidly establish whether the AV work is in the PUBLIC DOMAIN or in COPYRIGHT, it is the responsibility of the user to establish that a work is "ASSUMED to be in COPYRIGHT" and to go on with the workflow. (In ARROW, the protection term was always deduced by the system and the user could not intervene at such level.)

Degree of management of law national variations:

As mentioned previously ARROW interacted with the following main data source categories:

- TEL and VIAF – matching and clustering and Copyright status determination
- BIP (Books In Print) – retrieving commercial status information on the work manifestations
- RRO (Reproduction Rights Organization) – rights holders location and orphan status establishment

The ARROW workflow could be considered a workflow orchestrator. It handled national variations, workflow routing etc., but some "law pertinent" activities were delegated to external sources. In particular, we're referring to the last step of the process (Orphan status determination). In ARROW, it was delegated to the RRO to:

- establish rights holders (if publisher/author/heirs etc.),
- locate them, and
- declare if a work is orphan or not.

In FORWARD, the situation regarding this step is very different and more than a workflow orchestrator, the FORWARD workflow "embodies the law" since it has to act as an RRO itself by managing the establishment of rightholders (producer vs main authors) based on law, locating them and declaring if a work is orphan or not. FORWARD workflow complexity increases even more considering that there is a lot of national variation among member states. In other words, the FHI being the actor entitled to validate the Orphan status of an AV work implies that in FORWARD the FHIs have the double role of beginning a rights clearance request (as the Libraries in ARROW) and also to act as clearing centres themselves (as the RROs in ARROW).

Specific system requirements

In the case of ARROW for a variety of reasons there was no point in building an internal repository of parties, whereas the FORWARD project (according to D3.2) requires the system to be able to acquire 'intelligence', in other words to learn from previous queries. This implies that the system will progressively build its own internal database (e.g. of individuals and companies) in order to be able to use the information for future CoSAP and ADiSP. Obviously, the requirement introduces a certain degree of complexity in the FORWARD system.

As a consequence of the FORWARD workflow and its ability to enrich the FHI's data during the ADiSP process, the FORWARD system must be able to feed back such data to the participating institutions' database. In ARROW such a requirement did not exist.

During the ARROW and ARROW projects the OWD was not yet approved

When the first ARROW project began at the end of 2008 it was envisaged that the Orphan Works Directive would soon be in force. Unfortunately the OWD was published much later, the 25/10/2012 and ARROW and ARROW Plus projects defined the workflow on the basis of their own analysis. The sources of data needed to define the Copyright, Publishing and Orphan status were identified during the projects and included just the categories of external source defined by ARROW itself since those defined by law didn't exist. They mainly consists of these categories:

- National Libraries, The European Library and VIAF aggregators, constitute the sources of the bibliographic and contributors data and through them it is possible to infer automatically the Copyright status of a book.
- Books In Print provides the information on the availability of a Book and through them it is possible to infer automatically the Publishing status of a book.
- RROs establish rights holders, their location and declare if a work is orphan or not.

Interestingly, for ARROW the downside of not having the OWD in place became an opportunity as all the work done during the projects were recognized as highly valuable and in the directive the ARROW system is listed among the few sources to be used to perform a diligent search for published books.

2. METADATA ENRICHMENT

For the FORWARD system to function, the core set of metadata must be both available and in machine readable form. In Work Package 5 (WP5) the partners will enrich their metadata in order to provide the system with relevant information to perform the copyright status assessment process (CoSAP). To perform the copyright status assessment the system requires the work title, publication year, as well as the main contributors with their associated rights term information (birth date and death date).

Technical metadata requirements

The survey performed in D4.1 showed that most partners have relational databases with authority files, though some do not. For the partners that do have databases with authority files, it is expected that slight adjustments might be made to assist and streamline the mapping of the relevant metadata for delivery to the FORWARD system. However, it is foreseen that the system should also be able to maintain a centralized authority file database for companies and authors with related current copyright status/references.

While some partners may find resources to perform some technical adjustments to their databases, it is not directly part of the FORWARD project, and they are not required to do so. The system is expected to rely on a metadata package, which will presumably consist of an XML export from most partners, but should also be possible in a more manual mode for the partners not able to easily adapt their databases.

Metadata enrichment in WP5

The D4.1 survey showed that the main task for partner enrichment will be to research and register biographical metadata for the main cinematographic contributors. As WP5 leader DFI will perform surveys among the partners to monitor the progress of metadata enrichment on a number of key indicators, one of the main ones being the death date of the main contributors (authors), since this is essential to establish if a work is still in copyright.

Once the FORWARD system is up and functional, it will assist in performing the diligent search (ADiSP) for current location of main contributor rights, but a certain amount of manual enrichment, data entry and diligent search will still need to be performed in order to establish AV works as having orphan status.

The human effort in interpreting and declaring a AV work as orphan should not be underestimated.

The metadata enrichment has three main phases:

- **FHI local enrichment phase** – During this phase FHIs will enrich their local catalogues in order to complete information when missing or to normalize it in order to make it compliant with FORWARD. It is up to the FHIs to decide whether the data normalization will be persistent in their repositories or if it will be performed on the fly during the data export to FORWARD.
- **FORWARD enrichment phase for ADiSP** – Whenever an AV work is being processed, the FORWARD system will check if all information necessary for the whole workflow is present in the incoming FHI request file. If case of missing information the FORWARD system will provide the FHI the possibility to perform the pertinent diligent search via the FORWARD system. The gathered information will be used not only to carry on the workflow but will be also saved to enrich the original FHI record which can be exported back to the FHI in order to let them enrich their local catalogue.
- **FORWARD enrichment phase for CoSAP** – Besides enrichment of missing information, FORWARD can also export to the FHI the results of the CoSAP and Orphan status.

The details in specifying exactly which metadata will be hosted and retrieved from the partner databases, and which are more efficiently located and processed centrally by the FORWARD system still remain to be defined by FHIs during system design.

2.1 Metadata needed by the FORWARD system

In order to perform the FORWARD workflow a specific set of metadata is needed; this metadata set will enable the system to assess the rights status of an AV-work. As described in D3.1 metadata can be classified in three categories:

- cinematographic information,
- web identity search
- company information.

In general, cinematographic information is needed to identify an AV-work (title, country of origin, year of production, production company) and the main creators of the work (principal director, screenwriter, dialogue writer and composer). Web Identity Search and company information are needed to perform a Diligent Search to identify the rightholders of the AV-work

In this section we will analyse the specific metadata elements that are required for each and every node in the Decision Tree. To do so, we developed a matrix (see annex VI) that includes all the required metadata. Using this matrix we were able to develop an Entity Relationship Model (hereafter: ERD) that not only describes the metadata but also all the required relationships and field types (see 2.7 and 2.8 section}).

The first three columns of the matrix identify the locations of a specific element in the decision tree (column A), the relevancy of this element for the various Film Heritage Institutions (FHI) (column B), and the element itself that is described (column C). The entity is further subdivided into all the relevant attributes it contains. The attributes are identified by a label and a field type. In the case of an authority file, the attributes describe the fields of the authority file. For example, the contributors require an authority file with at least the attributes identifier, name, date of birth and date of death.

The matrix also contains a column describing the relevance of that element for the diligent search. We identified the type of search required (cinematographic, web identity search, company information) and the content type (film database, genealogical resource, etc). For specific information resources, we refer to D3.1. in which we described 367 possible databases that may be consulted to gather all the necessary metadata.

In the rows of the matrix, we identified the three stages of the example tree (applicability, copyright status and finding rightholders). These stages are refined with the specific steps of the tree (1.1., 1.2, etc). Furthermore, we mentioned the specific FHIs for which the element is relevant – according to their national legislation. In general, an element is applicable for all the participating countries, but as we mentioned in D4.1. there are national divergences. These divergences have been visualised in the matrix so that the system is able to deal with the specific provisions of national legislation.

In general, the basic set of elements is:

- The title of the movie
- Country of origin of the AV-work
- Production date
- Premiere / publication date
- Production company
- Principal director
- Screenwriter
- Dialogue writer

- Composer
- Producer
- Heirs
- Contracts

As we have seen these elements can have several attributes

- Names
- Dates of birth and death
- Documents
- Logical – Boolean (i.e. Yes – No)

These attributes have a certain field type; they can be authority files, logical, dates, hyperlinks and text attributes.

The system's output is the rights assessment of an AV-work. The system distinguishes between Copyright Status and Orphan Status:

- **Copyright Status**: can have one of the following values: "PUBLIC DOMAIN","UNDER COPYRIGHT", "ASSUMED COPYRIGHT", "UNDEFINED".
- **Orphan Status** – can have one of the following values: ORPHAN, PARTIALLY ORPHAN or NOT ORPHAN. The Orphan and Partial Orphan also indicate that the work in question is still in copyright.

2.2 Systematics of the metadata

On a logical level, an attribute can serve three purposes:

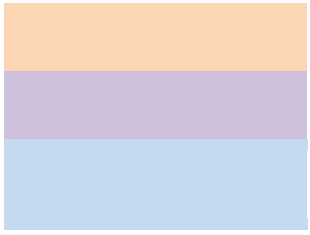
- It clarifies a step in the Decision Tree. This will support the rights assessment of the system
- It clarifies the search for rightsholders This will support the Diligent Search
- It serves both above purposes.

For example: the production date of an AV-work is relevant for the execution of the Decision Tree because it determines the rights protection term (> 70 year can determine that the rights status is Public Domain), but it is not as such relevant for the Diligent Search.

The place and date of birth of a principal director can be relevant for the Diligent Search (one knows when and where to look for his birth certificate) but it is not as such relevant for the execution of the Decision Tree.

Otherwise, the date of death of a principal director is relevant for the execution of the decision tree (once again, > 70 year can determine that the rights status is Public Domain) and for the Diligent Search (one knows when and where to look for death certificates, family announcements, etc).

We clarified these possible purposes with the following colours in the schema.



Required for the performance of a Diligent Search

Required for the execution of the Decision Tree

Information relevant for the execution of Decision Tree and the performance of a Diligent Search

Uncertainty of information is not addressed in this schema. We feel it would be best to choose a description method for uncertainty that adheres closely to the methods the partners already use, since this minimizes the effort needed by partners to provide the FORWARD system with this information. The exact manner in which partners describe uncertainty has not been investigated yet, and needs to be done in the scope of Task 4.3. An alternative choice is not to upload uncertain information, or apply a rule (so if the decade of creation is known, use the last year of the decade as creation date). This is a choice for each partner individually.

Once we had gathered all the necessary metadata to execute the Decision Tree and perform the Diligent Search we are able to outline an ERD (Entity Relationship Diagram) to describe the structure of the metadata.

2.3 Metadata: current and required

As described in D4.1 the FORWARD-system requires that the information will be harvested into a generalised data structure containing all the fields necessary to execute the legal decision trees of the different countries. We analysed the data structure according to completeness, consistency and usability of the data. In appendix we present an overview of the required metadata and their current presence, coverage and (if present) data structure of the FHI. This analysis gives an indication whether or not the current information resources of the FHIs are compatible with the FORWARD-system and what actions will have to take place during Work Package 5 to make interoperability possible.

The schema provides the main data of Appendix I, and compares them per FHI on data structure and coverage.

2.4 Unknown, inapplicable and unavailable values

Every element in the schema (except the title) can be non-existent for an AV work. But the semantics of this non-existence is unclear. It can mean one of three things:

- Unknown
  In this case the non-existent value indicates the field has been researched (by a diligent search) but could not be found. It essentially represents a conclusion of the diligent search, and means: the value is unavailable and will stay unavailable.
- Inapplicable
  In this case the non-existent value indicates the field is not applicable for this specific AV work, Person or Corporation. An easy example is Date of death: a Date of death is inapplicable for someone who is still alive.
- Unavailable
  In this case the non-existent value indicates the value is unavailable. But this is subject to change: a diligent search will change this value (either to unknown, or to the found value)

In D4.1 it was found that some partners in some cases maintain information that makes it possible for them to differentiate between Unknown and Unavailable, but the structure of this information needs further investigation. If in D4.3 it will be decided that the upload should contain a differentiation between unknown and unavailable the partners need to agree on the metadata definition for this kind of information.

Inapplicable is not specifically managed by the partners as far as we know, but is sometimes circumvented by using another field. To use the above example: an element Deceased indicating if a person is deceased solves the problem of the non-existent inapplicable Date of death.

2.5 Diligent Search

According to the metadata analysis (Appendix VI) the so called Assisted Diligent Search Process (ADiSP) is taking place during all the three segments of the Decision Tree (applicability, determining copyright status and finding

rightholders). A Diligent Search is here considered in the broadest sense of the word as a reasonable search to identify and locate rightholders and to enrich the metadata of an AV work in such a way that its copyright status assessment is possible by consulting all the relevant sources.

The user will upload a set of metadata to the FORWARD system on which the Copyright Status Assessment Process (CoSAP) will operate. Ideally this set of metadata is complete so rights assessment is possible, but in the case the information is insufficient the user has to be informed that an assessment is not possible yet. For example, when the date of death of the principal director is missing there can be two assumptions:

- the principal director is still alive or
- the date of death is unknown.

It is essential to know whether or not the user performed a Diligent Search on all the required metadata of the principal director. If s/he did so, the date of death is *unknown* after the search, which is relevant information in itself for the system: the AV work could be orphan. If the user didn't perform a Diligent Search yet, the date of death is unavailable, so he has to perform this search before the AV work can be assessed. This applies to all main contributors, and the producer or production company of the AV work. Generally the cinematographic resources will be sufficient to complete these biographical data but especially for the unfamiliar treasures of the film archives (amateur or family movies, documentaries) consultation of other resources has to be considered, such as civil registers and genealogical communities. Furthermore, there can be country specific rules that require a Diligent Search (such as the already mentioned extended protection term of main contributors that died for France during World War I and II).

In case the AV work is copyright protected a Diligent Search has to be performed to locate the rightsholders. It also has to be decided if additional information is relevant or not, such as employer relationship or state ownership. There may also be country specific rules in this segment of the system. The FORWARD system will supply a diversity of resources, because each MS will have to consult his own set of described resources according to their national law and to define the necessary depth of the Diligent Search. Whether or not the heirs or successors of the main contributors are part of the diligent search is a responsibility of the FHI. For the completeness of the schema heirs and successors are defined as such, so a diligent search tracking the transfer of rights is possible.

In general, an AV work will be assigned an orphan status if the current rightholders *cannot* be located after performing a Diligent Search. This is one of the data that are collected during the execution of the Decision Tree.

2.6 Overview Entity Relationship Diagram (ERD) schema

Figure 1 gives an overview of the complete ERD schema that clarifies the structure of the metadata described in the previous paragraphs. It is a conceptual schema, only used to describe the structure of information present in the system, and is not meant to be implemented directly.

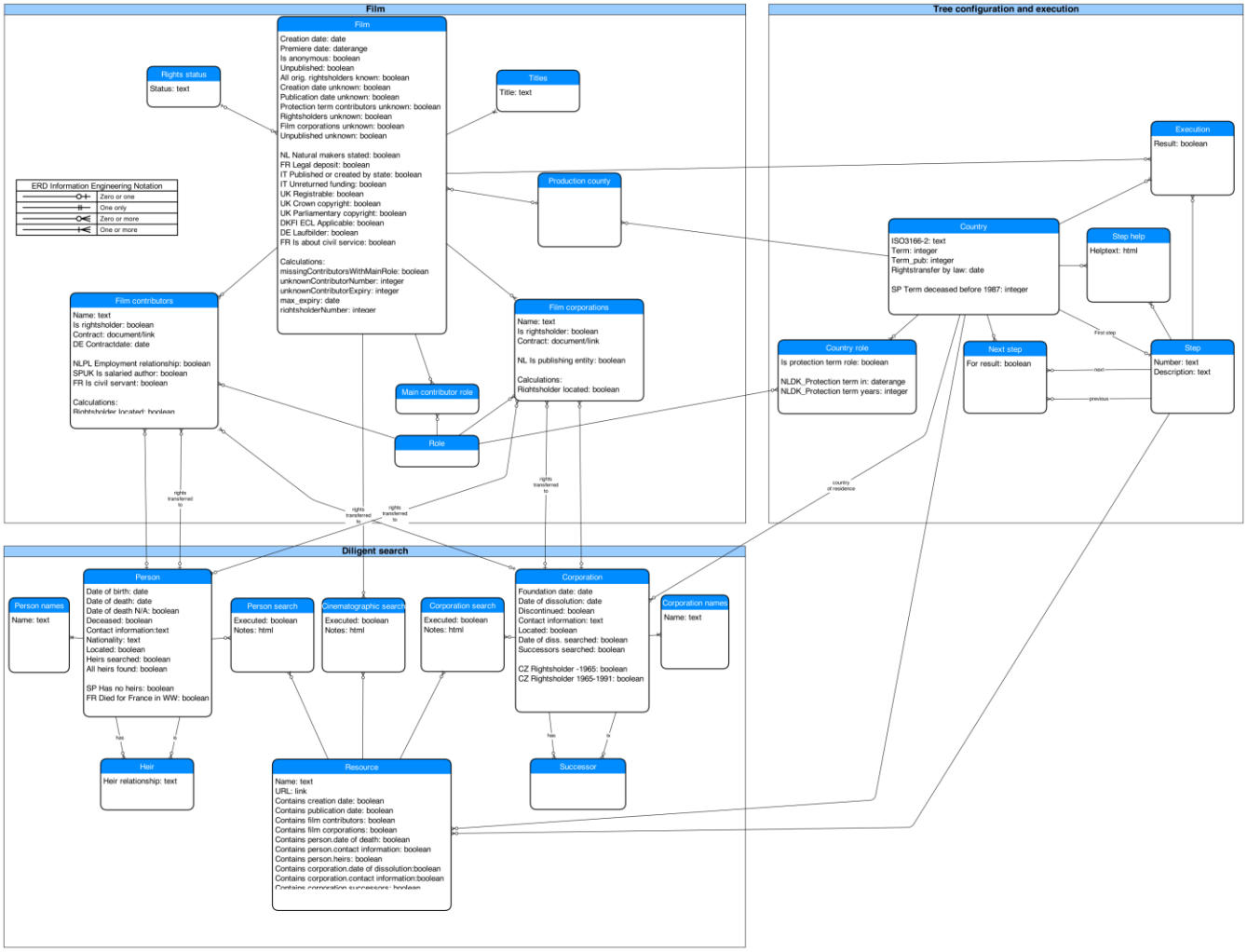


Figure 16 Overview ERD

The ERD consists of three sections:

- **AV** work
  Describing the filmographic information that results from the upload from the partner and information collected during execution of the decision tree.
- **Diligent search**
  Describing the results of the diligent search. The diligent search starts from what is known, i.e. the filmographic, person and corporation information in the upload from the partner. During the execution of the decision tree this information is enriched and completed where necessary as a result of the diligent search in resources specific for countries.
- **Tree configuration and execution**
  The decision tree consists of steps that differ per country. The decision tree of each country also requires different parameters.

"Country" is a very important entity: it influences the structure of the decision tree, determines the resources to be used in a diligent search, and is of course related to the filmographic information: the AV work was produced in a certain country.

As stated earlier, elements can be non-existent, and this can have one of three meanings: unavailable,

inapplicable or unknown. In the ERD this is modelled by:

- not making fields mandatory
- cardinality of relations
- introducing extra fields (such as Deceased, Heirs searched, Publication date searched, …)
- explicitly modelling the applicable Main contributors roles

The combination of an empty Publication date and Publication date searched is true represents Publication date unknown.

If a partner wants to treat their Unavailable fields as Unknown fields (i.e. the diligent search has been performed, but outside the FORWARD system), this should be part of the configuration for that partner used during the execution of his countries decision tree. So the FORWARD system only proposes a diligent search for fields that the partner wants to research. This needs to be further defined in D4.3.

2.7 ERD – AV work

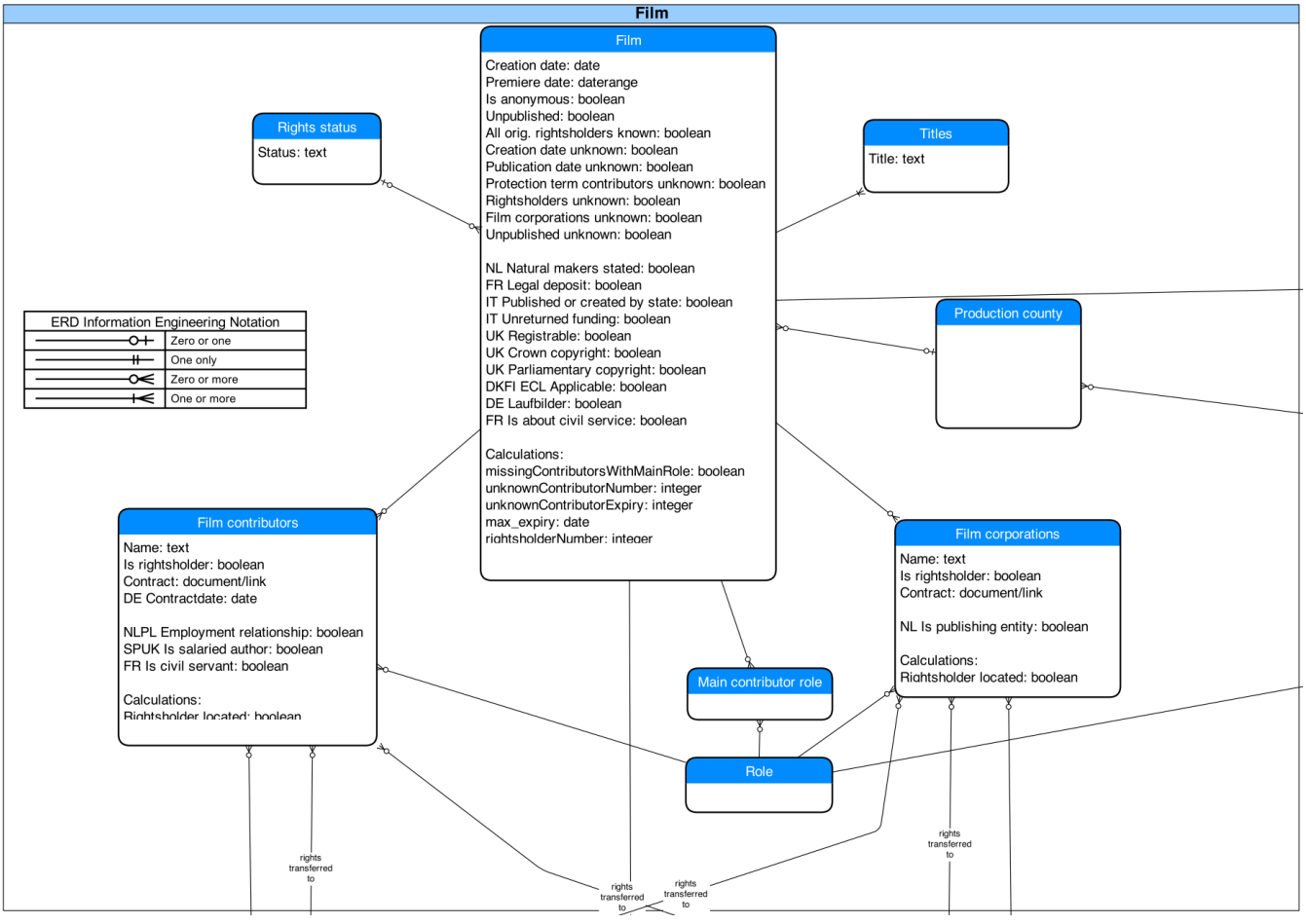


Figure 17 AV work section

2.7.1 Uploaded data

The uploaded data can contain the following fields (not all the countries will upload all the data):

- Creation date

- Publication date
- Is anonymous
  If it is known beforehand that the work has been published anonymously.
- Unpublished
- Titles
- AV work contributors
  Persons that contributed to the work in a given role. The upload includes "persons" for the partners that use an authority file.
- AV work corporations
  Corporations that played a role in creating the work. One of the most important corporations is the production company. The execution of the decision tree is determined by the country of residence of the production company. The upload includes "corporations" for the partners that use an authority file.
- Production country
  If the country of residence of the production company is unknown, often the production country is present in the data, and will be used to determine the decision tree execution.

2.7.2 Data collected during execution of the decision tree

We will now present how the ERD schema relates to the decision tree execution. Chapter 1 depicts the accurate and complete logic of the decision tree.

The first step is to determine the required country for which the decision tree is executed . The logic looks at *Film.Film corporations[Role=Production company].Corporation.Country of residence.Country*, if that information is not present *Film.Production country.Country* is used. In case of multiple choices or missing information the procedure is not yet clear.

To determine the copyright status it is essential to know which roles need to be considered. For example in a silent movie the role Composer is not applicable, so a diligent search for the composer is not necessary. This is modelled in the entity *Main contributor role* (the 'default' roles for a country are part of the Tree configuration section). During a diligent search it is also possible to identify whether the AV work has been published anonymously, which can then be stated in the field *Is anonymous*.

For partners that are not able to create an authority file, the upload only contains the *name* of the *Film contributors* and *Film corporations*. The first diligent search that has to be executed is the diligent search in *Persons* and *Corporations*, so the *Film contributors* and *Film corporations* are related to *Persons* and *Corporations* with their data.

The entity *Film* also contains fields that reflect actions that have been executed (i.e. diligent searches that have been executed). This is necessary because an unknown publication date that has been searched for but not found has a different status in the decision tree than an unknown publication date that has not been searched yet.

The field *Film.All orig. rightsholders known* is used to distinguish cases where only some of the original rightsholders are known (even after a diligent search), so even if all the rightsholders known to the system have been found, the *Film* is still a (partial) orphan.

The entities *Film contributors* and *Film corporations* are boolean fields to indicate if the person or corporation was a rightholder for that film. These persons and corporations are the starting point for the search of rightholders. The possible rightholders roles are modelled in the tree configuration section.

The entities *Film contributors* and *Film corporations* contain a calculated field that indicates if the rightholder or

their heir has been located in Step 3: Rightholders determination of the decision tree.

The entity *Film* also contains some calculated fields used by the decision tree in determining the copyright status.

The entities *Film*, *Film contributors* and *Film corporations* also contain country specific fields (since each country has different legislation and its own resulting decision tree). These fields are prefixed with the two letter ISO code of that country. Based upon the results in Deliverable 4.1 some fields are already defined. Note: since the partners themselves are responsible for their decision tree this is only indicative of the necessary fields specific to the decision tree of each country.

2.8 ERD – Diligent search

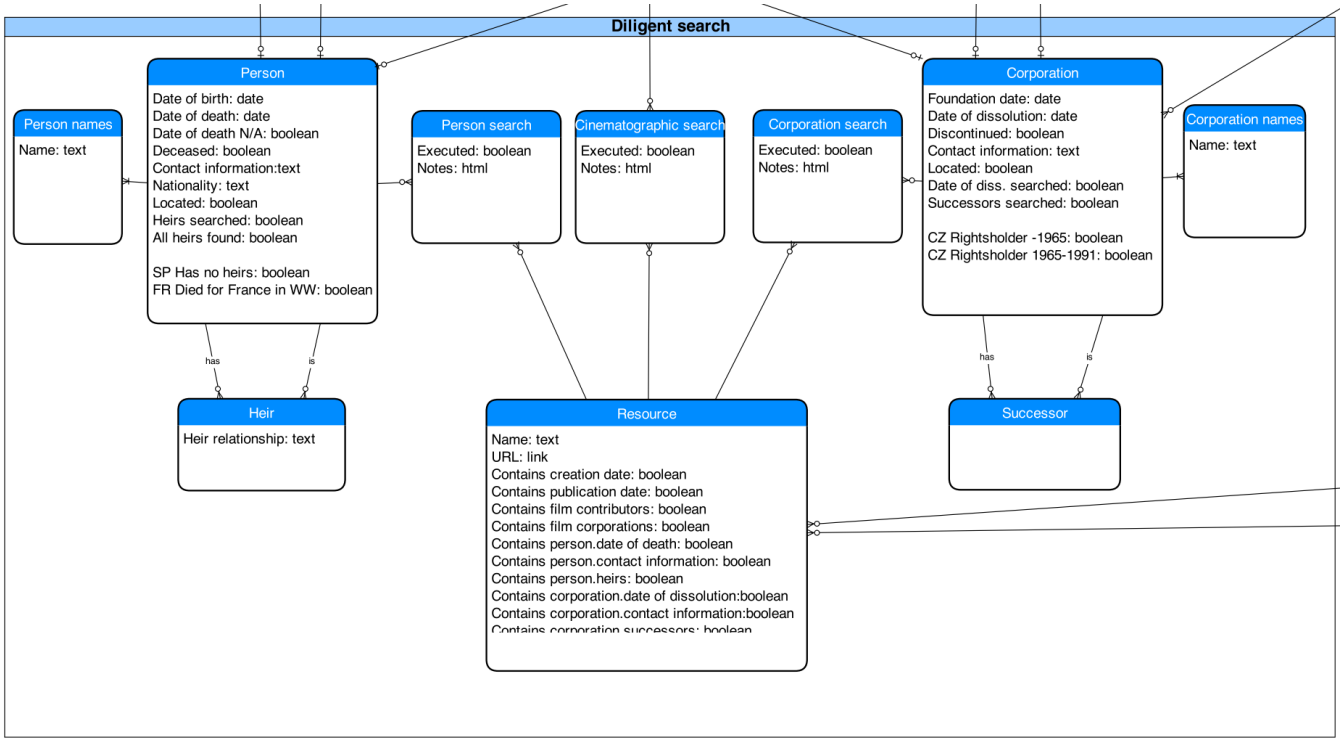


Figure 18 Diligent search section

2.8.1 Uploaded data

As part of the upload of filmographic data, person and corporation data is also uploaded (and related to the film through the entities *Film contributors* and *Film corporations*), where they are present in the database of the partner.

For the entity *Person* the following fields can be contained in the upload:

- Date of birth
- Date of death
- Deceased
  Field indicating whether the person is deceased. This is necessary because absence of Date of death can

mean the person is alive or that we do not know the date of death. This field translates into Date of death N/A.

- Contact information
  Fields necessary in case a rightholder is located, and can be contacted through mail, telephone or other means.
- Nationality
  This field is necessary for the selection of the appropriate diligent search resources. Searching for the date of death of a French individual that worked on a German AV work will require French resources.
- Heirs

For the entity *Corporation,* the following fields can be contained in the upload:

- Foundation date
- Date of dissolution
- Discontinued
  The 'deceased' field for a corporation
- Successors
- Contact information
  Fields necessary in case a rightholder is located, and can be contacted through mail, telephone or other means.

2.8.2 Data collected during execution of the decision tree

We will now present how the ERD schema relates to the decision tree execution. Chapter 1 depicts the accurate and complete logic of the decision tree.

In Step 2: Copyright Status of the decision tree, a diligent search is executed for the 'Main contributors' (the contributors that are relevant for determining the copyright status). It is therefore possible that "Persons" as AV work contributors and "Corporations" as AV work corporations are added to the system. The fields *Film.missingContributorsWithMainRole*, Film.unknownContributorNumber and *Film.unknownContributorExpiry* and *Film.max_expiry* reflect the results of the diligent search. The entity *Cinematographic search* is used to store information on which *Resources* have been consulted.

For all "Main contributors" a diligent search for the *Date of death* is performed. If the *Date of death* is unknown and *Date of death N/A* is not true it means the date of death is not found after executing the diligent search. The entity *Person search* is used to store information on which *Resources* have been consulted for the diligent search.

In Step 3: Rightholders determination a diligent search is performed to locate the current rightsholders, starting from the original rightsholders. This can include a search for the *Heirs* or *Successors*; this means adding new persons and corporations to the system, and enriching the persons and corporations data. The field *Located* indicates that a rightholder has been located; this means that a living person or existing corporation can be contacted using the contact information. The fields *Date of death searched*, *Date of dissolution searched*, *Heirs searched* and *Successors searched* are set to true once the diligent search has been executed. The entities *Person search* and *Corporation search* is used to store information on which *Resources* have been consulted for the diligent search.

A *Resource* can be consulted during a step in the decision tree for a specific country. For each resource the kind of information available in the resource is indicated, so that the user can make an informed selection of the resources to consider during the diligent search.

Person and *Corporation* also contain some specific fields for specific countries. In the Czech Republic the rights

of all AV works created before 1992 reside with the National Film Archive or a State Fund depending on the creation date. These representatives are added as corporations; the fields indicate which of these representatives own the rights, In Spain the rights to a AV work belong to the state if no heirs can be found.

2.9 ERD – Tree configuration and execution

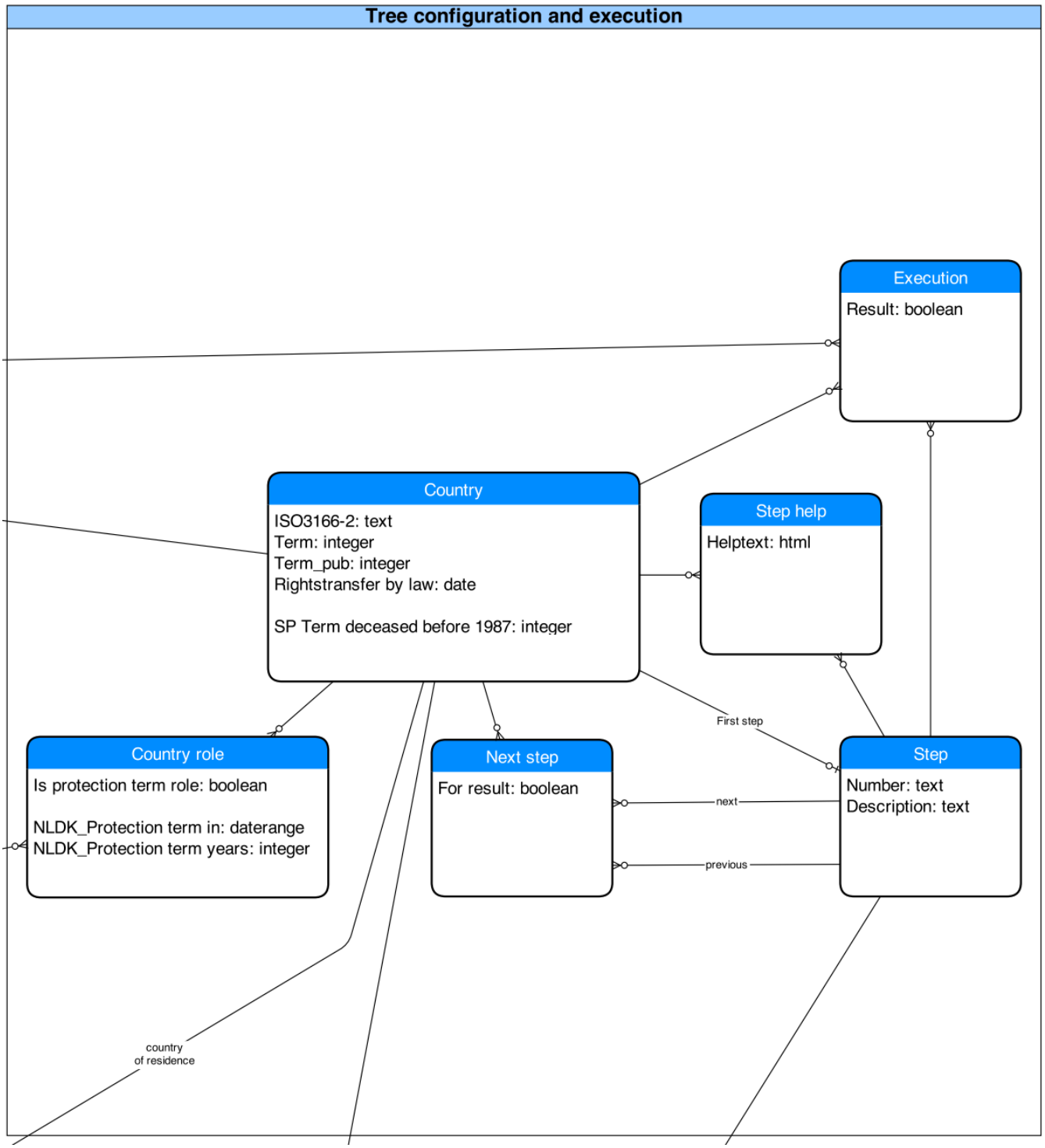


Figure 19 Tree and country configuration section

This part of the ERD depicts a conceptual structure of the decision tree, including possible adaptations for each country. We use this structure to indicate where possible adaptations for the specific countries can be found, and how the decision tree can relate to the metadata. It is in no way a guideline for implementation of the decision tree; those can be found in Chapter 1. The parameters that are basic input for the decision tree of a specific country are depicted as fields of the entity *Country*. The entity *Country role* describes which roles are potential copyright status roles (roles used to determine the copyright status). For instance in the Netherlands and Denmark, Transition Law needs to be taken into account when determining the roles relevant to the assessment of copyright status. The additional fields here are meant as an example of metadata model adaptation.

The decision tree itself consists of steps. Each step can consist of:

- a function that calculates the outcome of that step for an AV work.
  the outcome is “true” (yes), “false” (no) or “null” (in case data is missing, and a diligent search needs to be executed)
  The outcome of a *Step* for a *Film* is administrated in the entity *Execution*.
- a function that retrieves the information that is incomplete for this step (for which the program that calculates the outcome returns “null”), so the user can complete this information
- a list of resources (per country) that should be consulted to complete the “incomplete” information
- Per country: a next step for the outcome “true” and a next step for the outcome “false”
- Per country: a help text so the user can make an informed choice about how to complete information.

A country has a first step: this is the entry point into the decision tree for that specific country.

If a step generates a result, it is not necessary to bother the user; the next step can be retrieved for the country, and the program for the result of that step can be executed.

3. CONCLUSIONS AND NEXT STEPS

3.1 Conclusions

Deliverable 4.2 focuses on the analysis and design of FORWARD workflows to assess rights status(es) of AV works in the countries represented by the project partners.

The legal analysis conducted in the previous deliverables shows that protection term and rightsholder determination are not completely harmonised across EU Members States. In particular, the level of harmonisation regarding the rightsholders of AV works is limited and variations between the different countries emerged during the workflow design. This implies that the FORWARD workflow has to address 28 MS with legislations that only to some extent are consistent.

The lack of comprehensive, transnational databases imposes the use of a plethora of smaller, national (and even regional, local, personal) databases; the implementation of the FORWARD workflow in the different member states includes quite an extensive list of resources to be queried that can be considered mandatory.

The analysis of Available Resources in the AV sector revealed a fragmented and not harmonised nature of the AV work domain data. There are no widely used standard export metadata schemas in the AV sector, FHI's Authority files in most of the cases do not contain standard identifiers of the entities they represent and FHI's databases do not always contain all the information that is necessary to carry out the "*assisted diligent search process*" (ADiSP). Even when the information is present, there may be many differences in format and types among FHIs.

This brought forth the requirement that the FORWARD system be more like a diagnostic system, where it is possible to perform queries in order to identify the rights status and rightholders of the AV works, but where it is the FORWARD workflow that guides the process, and the diagnosis is conducted throught the ADiSP.

Based on these analyses and on the System requirements the FORWARD workflow has been designed to be configurable, flexible, standard-oriented, reusable and scalable:

- **Configurable**: in order to manage the variations of national law and operating procedure of each country, the FORWARD system will define a workflow instance for every country. The instance will be set up based on the workflow configuration settings provided for each country by the partner in charge. The workflow can be configured at two different levels:
  - Workflow branches to be applied or not
  - Parameters for each selected branch to be applied or not

  The workflow presented in this document was approved by the Dutch, German and Finnish archives, countries where a national Orphan Works Law is already implemented at the time of completing this document. For the other countries of the partnership, it was agreed that the configuration of their workflow will be completed during the course of the project, as soon as legislation will be implemented.

- **Flexible**: this characteristic can be found all across the FORWARD system, which is designed for easy workflow modifications and extensions in order to adapt to different national scenarios.

  The FORWARD workflow contains processes to establish the value of elements that are necessary for determining the rights status of the work. The fact that the Diligent Search complexity can vary is also an evidence of workflow flexibility to handle different national scenarios. For example, there can be countries that provide the anonymity value of an AV work in the metadata and there maybe countries that need to perform an external diligent search.

- **Standards-oriented**: to overcome the lack of standards in use among the FHIs, FORWARD will define an export schema to export FHI's data for processing in FORWARD. FHIs will be responsible for correctly exporting their data in the established schema.
- **Reusable**: The FORWARD system establishes a parametric diligent search engine that provides reusability in the workflow in case diligent searches are necessary for additional elements of the workflow.
- **Scalable**: the workflow has the capability to handle further workflows, i.e. other countries or further data sources.

In synthesis the FORWARD workflow is designed in accordance with the national legislation of each of the FORWARD Partners. One of the main features is that it cannot be automatic (as in ARROW), but human intervention is required in the following cases:

- In many – or most – cases the user has to validate the search results retrieved from automatic querying of external sources. In case no result is obtained from the automatic sources, the user has to manually perform searches in the external sources which do not have an automatic interaction in FORWARD and store the results obtained.
- The user has to interpret the data (or lack thereof) and make a decision. This does not only pertain to a specific search result, but also to the interpretation of a set of results
- Only Beneficiary Users as defined in the Directive can declare a work Orphan and the FORWARD workflow has to assist them in this manual assessment.

In addition to the design of the FORWARD workflow, this document presents also the analysis and definition of the existing metadata structures required for the execution of the FORWARD. Based on this, three enrichment phases have been identified:

- FHI local enrichment phase – During this phase FHIs will enrich their local catalogues in order to complete information when missing or to normalize it in order to make it compliant with FORWARD. It is up to the FHIs to decide whether the data normalization will be persistent in their repositories or if it will be performed on the fly during the data export to FORWARD.
- FORWARD enrichment phase for ADiSP – Whenever an AV work is being processed, the FORWARD system will check if all information necessary for the whole workflow is present in the incoming FHI request file. If case of missing information the FORWARD system will provide the FHI the possibility to perform the pertinent diligent search via the FORWARD system. The gathered information will be not only used to carry on the workflow but will also be saved to enrich the original FHI record which can be exported back to the FHI in order to let them enrich their local catalogue.
- FORWARD enrichment phase for CoSAP – Besides enrichment of missing information, FORWARD can also export to the FHI the results of the CoSAP and Orphan status.

3.2 Next steps

During the coming three months leading to the completion of D4.3, two main tasks will be achieved.

The first is under the responsibility of CINECA team that will be involved in the System Design, where the architecture, components, modules, interfaces, and data of the FORWARD system will be defined. CINECA and the FHI partners will also be involved in clarifying some open issues like:

- the way Beneficiary Users will send the data to the FORWARD system: through an initial load or through iterative submissions. During the design both alternatives will be evaluated further and the most

appropriate strategy will be selected for the FORWARD system.

- the workflow configuration parameters will be provided and validated by each FORWARD country and each workflow has to reflect the national legislations of the 10 FHI FORWARD partners.
- a first draft of a common metadata schema has to be drafted in order to help the harmonisation of the data coming from the FHIs before being processed in FORWARD.

In parallel with the design, the metadata enrichment at FHI local archive level will need to be detailed and will start. This process may require not only the integration of missing information but also some data normalization.

LIST OF REFERENCES

ANNEX I: Template Workflow Configuration Parameters

ANNEX II: German Workflow Configuration Parameters

ANNEX III: Finland Workflow Configuration Parameters

ANNEX IV: Netherlands Workflow Configuration Parameters

ANNEX V: Danish Workflow Configuration Parameters

ANNEX VI: FORWARD workflow metadata analysis

ANNEX VII: Workflow metadata comparison per country

ANNEX VIII: Glossary