FORWARD

Project Number: **325135**

Project Acronym: **F O R W A R D**

Project Title: **Framework for a EU-wide Audiovisual Orphan Works Registry**

Instrument: Pilot B

Thematic Priority: CIP-ICT-PSP-2012-6

D3.2 System requirements

Due Date:	31/10/2014
Submission Date:	11/11/2014
Start Date of Project:	01/07/2013
Duration of Project:	36 months
Partner in Charge of Deliverable	CRB
Version Status	Final
Dissemination Level	PU[1]
File Name:	D3_2 FINAL.pdf

---

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

Revision History

Revision	Date	Author	Organisation	Description
N°1	22/09/2014	G.Scipione	CINECA	Draft
N°1	16/10/2014	NMazzanti	CRB	v.1
N°2	20/10/2014	G. Scipione	CINECA	v.2
N°3	28/10/2014	NMazzanti	CRB	v.3
N°4	10/11/2014	NMazzanti	CRB	FINAL

Table of Contents

EXECUTIVE SUMMARY

The first phase of the FORWARD project consisted in assessing the legal and operational environment in which the system is to operate in order to define its business requirements and functionalities.

Furthermore, this first phase highlighted the differences existing among MS's legal systems, OW legislation (when already in place), informational resources available, and practices.

This first phase allowed the definition of the FORWARD system's requirements and workflow. These are described in the present document D3.2, and in D4.2.

This document also describes the key User groups, as well as the 'products' and 'by-products' of the system, on which a long term sustainability is based.

Requirements are broken down for different process (copyright assessment, diligent search) and relations with external and internal sources are defined at high level, as well as interactions with external users (like the OHIM database)

Key requirements obviously include performance of a diligent search and recording every step of it, so that a precise record of the search can be retained by the beneficiary (within the system and/or outside of it).

Also, it is clear that the FORWARD system will operate based on sets of rules that are specific to the Member State in which the diligent search must be performed. These rules will be kept and maintained by the system with the support of national FHI(s). This is both a functional and legal requirement according to the OWD.

A glossary of the terms used within the system is also established.

PURPOSE OF THE DOCUMENT

Overall goal of WP3's was twofold:

1. **Define the environment in which the FORWARD system is to operate**, by analysing legislations at MS-level, operational conditions, users (and/or customers), available (and not available) resources, metadata and interoperability etc.
2. **Define the high-level requirements of the FORWARD system** so that they can drive the system design phase.

The first objective (Task 1) was achieved in previous Deliverables (most importantly in D3.1)

The present document concludes Task 2 (and consequently the work in WP3) by defining, at high functional level, the functionalities and the requirements of the FORWARD system. These are conceptually grouped in three broad areas:

- Definition of users and of their interactions with the system
- Definition of the "products" of the system (as per services, results, both as primary and as 'by-products')
- Definition of the system's use of informational resources (databases and alia)

All of the above does also define and describe the workflow to assess AV rights (as a product and as result of the system elaborating a given set of resources).

RELATION WITH OTHER DOCUMENTS

Following the definition contained in the paragraph above, this document derive from, and further elaborates the findings contained in previous deliverables D3.1 and D4.1.

Also, this document is the basis for D4.2 and D4.3, which will describe how the system is to be technically designed in order to meet the requirements set out in this Deliverable.

This document also constitutes the achievement of the first Milestone of the project "MS1-High level functions decided".

1. BUSINESS REQUIREMENTS

1.1   Introduction

The term "business requirements" is used in this document in the meaning it is used in the IT service environment as "functionalities required" to meet the demands of the users; it refers to user needs or user requirements, describes what users do with the system, such as the activities that users must be able to perform.

The business requirements here presented will be broken in the next chapter in more detailed functional requirements, which outline what needs to be delivered.

The starting point

The FORWARD project was initiated under a Call from the Commission to design and implement a system capable of achieving goals similar to those achieved by the ARROW system in the book sector.

The original remit (the Call) included the creation of a Registry of Orphan works and a network of "clearing centres" able to identify Orphan and "Out-of-Commerce" works.

Many factors intervened since the publication of the Call and the writing of this project that influenced the focus of FORWARD:

-   The Orphan Works Directive (OWD) was approved and it is being implemented literally while this Deliverable is written

-   The OWD mandates the creation of a free, open to all Registry of Orphan Works to be managed by OHIM (Office for the Harmonization of the Internal Market)

-   No consensus could be achieved among stakeholders about the concept of "Out-of-Commerce" in the field of AV (with positions ranging from "nothing is out of commerce in AV field" to "almost everything is") consequently, the whole issue is out of the scope of FORWARD.

Several basic considerations (obvious to those who deal with AV works and their rights but not necessarily widely known) also inform the Requirements:

-   Unlike in the book sector, no exhaustive databases of AV works under copyright exist; i.e. no sources similar to the in-Print books catalogue;

-   Information about authors contained in the databases of the CMOs (Collective Management Organizations) are also largely incomplete; first of all because in many MS no specific CMO for AV authors exists, secondly because even when they do, they largely cover authors involved in TV productions – i.e. they hardly cover authors active prior to the 60s

-   As a consequence, other sources need to be harvested – these are mainly the databases hosted

by the FHI in each MS

- These information resources must be completed with information about the actual rights owners; in the AV sector one must take into due consideration the role of the "producer", which is way more complex and critical than that of the publisher in the book sector

- This, coupled with the above considerations, impose that a wide range of resources are queried before ascertaining the rights situation of a given work (e.g. they might require a check at the Chamber of Commerce to ascertain if a given company is still active).

- Because of lack of harmonization, we are effectively faced with 28 MS with legislations that are consistent only to some extent

- The OWD indicates that the diligent search must be carried out in the Member State of the work's producer's headquarters or habitual residence; this means that diligent searches must be MS-based, following local legislations and usually procedures defined in each MS (e.g. in terms of which resources to use)

- The European AV sector is (and has always been) one of co-productions and of authors working across borders. This means that works can be considered being produced in more than one country, and the authors can be resident (then or now) in different countries from the one where the work was produced. This leads to a high percentage of cross-border diligent searches

- Only "beneficiaries" as defined in the Directive can declare a work Orphan (following a diligent search); these are public mission institution who actually hold the work (or a copy thereof) in their collections

- Obviously only a work that is in-copyright can be declared Orphan

- Works that are declared Orphans must be listed in the OHIM database (i.e. information must be transferred towards it; the act of recording the info in the OHIM database has no bearing on the status of the work as Orphan

- Beneficiaries must retain a precise record of the Diligent Search they carried out on each work

The above considerations and factors concur to define the required functionalities of the system (business requirements).

Differences against the ARROW system

The key factors and characteristics of the AV environment and their impact on the OW field are also useful to understand the main differences with the ARROW system, which we recall here, in an extremely shortened and simplified way:

The main challenge in the ARROW system was the identification of the Work against its Expressions and Manifestations (e.g. different translations, editions,…). This can be achieved by an automatic procedure of clustering (human intervention might occasionally be required).

After that, the Author of the Work was identified and sought for in various databases to verify his/her date of death to make sure the work is in-Copyright, and if this is the case, the name and other related

information are checked against publisher's and CMOs databases to define who can clear the work's rights.

The main informational sources are external to the system (VIAF, TEL, in-Print books, etc.); they are highly organised and comprehensive, and relatively small in number (thanks to a concentration/federation of local databases, like the ones from the publishers).

As a result, while allowing for some human intervention at critical junctions (e.g. validating the clustering phase), the ARROW system is highly automatic and is designed to carry out queries and produce answers in a fully automated way.

The picture is profoundly different 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. This is of course a simplification in the process. Unfortunately, it is the only one.

- In case of AV works, there are many authors (and the list is not necessarily the same in all MS): at least four, so the diligent search must be repeated *n*-times.

- The "producer" (either a physical person or a company/body) play a key role as rights aggregator; this role depends both on contractual terms between the producer and the authors, and on legislation (e.g. this can or not assume a transfer of rights to the producer), or on both.

- 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 OWD in the different MS 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.

- As mentioned earlier, and differently from the book sector, the European AV sector is (and has always been) one of co-productions and of authors working across borders.

- These factors mean that the search that FORWARD must perform is much more complex (multiple authors, multiple countries of production…) than the one in ARROW. Plus, it must be tailored on national legislations and sources.

- Also, in many – or most – cases, the search will require human intervention. First of all because the system might require information that cannot be found in the resources that can be queried automatically, and secondly because some level of human interpretation of the data (or lack thereof) is likely to be required.

- Last but not least, the act of declaring a work Orphan implies a certain level of legal responsibility (and consequent liability) that must be the result of a conscious act of the beneficiary's representative. In other words, a final, human validation step is practically unavoidable (this can also take the form of a 'trust the machine' automatic validation – still, it must be based on an

overarching, rule-based human decision).

FORWARD's products and by-products

Obviously, the introduction of a freely accessible database at OHIM affects negatively the possibility for FORWARD to offer a fee-based service regarding OW to a wider public.

The fact that the OWD does not authorise commercial use of OW, and that the only beneficiaries are public-mission, non-profit institutions also limits the potential number of beneficiaries (which obviously includes the project's partners).

Having said that, the number of beneficiaries is relatively high for a B2B service like FORWARD, and if FORWARD manages to simplify trans-border diligent searches, users can amount to dozens per MS.

Somehow similar to ARROW are the 'by-products' of the system. These include the identification of works that are in the Public Domain (and/or of a date when they will fall into the PD), and the identification (and possibly location) of rights holders, in the case of non-Orphan works. These "by-products" offer in fact more chances to be commercially interesting to a wider number of customers – mostly commercial archives and users who need to ascertain the copyright status of a given work.

The high complexity of the searches, the level of interactivity with the searcher, the confidentiality of some of the (sources of) information used by the system, lead to the fact that it is very likely that FORWARD will provide "expertise services", rather than allowing 'customers' to directly perform searches. In other words, external users will have to send requests to FORWARD (as such, or via one of the clearing centres) that will analyse the information, run the search, and report back the results.

This model would also correspond to the Call's demand for 'clearing centres', and has the advantage of having actually a higher added value (= price).

1.2   Basic requirements

FORWARD's core requirements can be easily defined by factoring in the original remit, the DoW, the abovementioned considerations, and the findings of the previous Deliverables.

The core requirements include:

1. The system supports the carrying out of diligent searches necessary to declare a Work Orphan. We say that the system "supports" vs. "carries out" the diligent search because the system in itself does not *declare* a Work Orphan, and human intervention to do so. Human intervention is very likely to be necessary at one point or another anyway.

2. As only a representative of a *beneficiary* (as per the OWD and national legislation) can declare a Work Orphan, System's administration (and FORWARD governance) will include a procedure to officially authenticate an entity as an beneficiary.

3. Each beneficiary will be able to open several active accounts, according to preset profiles (see below), in order to interact with the system.

4. Internal users (i.e. partners of the project) are beneficiaries by default.

5. Beneficiaries' active FORWARD accounts are given one or more set(s) of privileges. Each

beneficiary (*Beneficiary User* - BU) can assign one or more of the following privileges to one or more accounts depending on its internal organization:

   a. Initiate Query – can submit, individually or in bulk, queries to the system
   b. Input data to Query - interacts with system (e.g. feeding requested info),
   c. Manage Queries – manages queries, can abort query(ies) (e.g. when information is insufficient)
   d. Status validation – validates copyright status, declares Work as Orphan
   e. Read process – has access to all processes, cannot interact (e.g. to evaluate righsholders' claims)
   f. Read results – has access to results after process completed (validation), cannot interact

6. A query will always be initiated by submitting a request structured according to a standardized procedure.

7. This procedure will include a defined set of metadata, organized according to a defined schema (to be defined in D4.3). Any work necessary to identify the work will be carried out prior to the submission; it is obviously possible that the research does not lead to the identification of the Work; in this case, procedures for submitting unidentified works are foreseen.

8. Internal Users will have the option to submit also works that are clearly Orphans (e.g. published works without author's or producer's names) if they want to exploit the ability of the system to act as a database of diligent searches. Alternatively, they can opt to handle these queries outside the system and record the diligent search manually

9. It is considered advisable that traces of such manual searches are kept in the system, for reasons of completeness,

10. The set of metadata input by the user is always first matched against the System's internal databases. If a work matching the data is found, this is reported to the user along with the work's copyright and orphan status information, and the process ends. Partial matches are foreseen to require user intervention.

11. It is also possible that by matching two or more records of the same Work that are present in the system, more information on the Work is found (e.g. authors name present in record A, while absent in record B at the moment of ingestion). In this case the User is informed and/or metadata automatically updated.

12. In order to initiate an *"assisted diligent search process"* (ADiSP), it is necessary to assess if the work is still in-Copyright and not in the Public Domain (*Copyright status assessment process* – CoSAP)

13. CoSAP is obviously based on rules that are adapted to national legislation (MS-Law)

14. If CoSAP finds that the work is in the Public Domain, the result is communicated and the process ends there

15. If CoSAP's results indicate that the work is in-Copyright, the process automatically moves on to initiate an ADiSP. In order to allow the system to proceed also when the in-Copyright status cannot be ascertained with 100% certainty, the system foresees the option of "Assumed in Copyright" that allows to proceed to ADiSP.

16. In order to carry out all processes (CoSAP, ADiSP) the system processes information resources that are internal and external. Internal resources are those belonging to the partners (under the provisions contained in the Consortium agreement)

17. Internal users (partners) constitute the '*clearing centres*' as defined in the Call and in the DoW, as they can process requests also on behalf of others (see below, Users sections)

18. The diligent searches will be carried out according to the legislation, rules and regulations of the relevant MS for any given work ("*Member States Diligent Search Rules*"- MS-DSR)

19. As the decision of the relevant MS is in itself a matter of human judgment and interpretation, the request submitted to the system includes already this information (i.e. the system does *not* define the applicable MS).

20. If multiple MS are relevant (e.g. co-productions, cross-borders authors,..) the system is able to repeat the procedure according to MS-Law and MS-rules of *n*-MS.

21. The system interfaces with users internal and external to the system (see below in Users sections).

22. The OHIM database is an external user that is already built-in into the system.

23. The system interfaces with users via messaging whenever a human intervention is required. In the case of a 'possible Orphan', this will happen *at least* once, at the end of the process in order to validate the Orphan status.

24. The system will log, record and maintain all the steps of the diligent search, both those carried out by the system and those due to human intervention

25. This implies that a specific interface is designed so that information can be entered in the system. In other words, not only the result of the each step is recorded ("*the production company was dissolved in 1963*") but also the how ("*User XYZ knows this by: checking the Chamber of Commerce records on XXXXX date*").

26. The records of the search logs are stored by the FORWARD system, and are accessible only to the beneficiary who carried out the search. Alternatively, they can be downloaded locally and erased from the system, if the beneficiary so wishes.

27. Once a work is declared Orphan (after validation of the beneficiary), the information is transferred to the OHIM database. The modalities of this transfer are defined by the beneficiary according to its national legislation (they go to OHIM directly or via, in parallel to, prior to, etc. a third body)

28. The information about a given work being declared Orphan is also retained and made available by the System according to rules defined by the relevant beneficiary and by the FORWARD governance.

29. Prior to starting a diligent search, the system checks whether the work is in-Copyright or in the Public Domain via the CoSAP and according to rules defined for each MS (MS-Law).

30. If the work is found to be in the Public Domain, the beneficiary is informed of the result. The search

ends there. Records of the search are stored by the system or delivered according to the rules set up by the beneficiary. The information is also transferred to a registry of Public Domain works that is managed by the System. Access to the list is granted according to rules decided by the FORWARD governance (e.g. free to all partners, for-a-fee for others…)

31. In case the work is found to be in-Copyright, the *"assisted diligent search process"* (ADiSP) is initiated, according to *"search-rules"* defined per each relevant MS (MS-DSR)

32. Independently from the actual technical solution implemented, in terms of system's logic the ADiSP is performed strictly and uniquely according to a silos approach. In other words the system treats each ADiSP within the boundaries of one set of diligent search rules applicable to one relevant MS as required by the Directive and by local legislation (MS-DSR). This is needed because for the ADiSP results to be valid, the system must be able to demonstrate that the search is performed according to the relevant MS-DSR (or of n-MS-DSR for co-productions or contributors living in different countries, where *n* is the number of relevant countries).

33. To that purpose, the system must maintain (i.e. keep and update whenever necessary) as many sets of MS-DSR and MS-Law as necessary (one per MS activated in the system).

34. Each MS-DSR describes the logical processes necessary to the ADiSP, and the sources of information to be queried, being them internal or external to the system

35. The sources of information are defined as "internal" (*Internal Metadata Source* - IMDS) when they derive from resources owned and made available by the FORWARD partners. At present these are limited to the project's partners, but they are expected to grow.

36. External sources of information (*External Metadata Source* - EMDS) include all other sources. In the system, these are categorized according to their relevance and to their "legal value" (i.e if they are mentioned in the relevant legislation as 'mandatory').

37. Deliverable 4.3 will define in details the metadata structure for communications between the system and IMDS and EMDS.

38. Considering the number of sources that might be necessary and the technical and/or procedural issues in accessing them (i.e. the system might be authorized to access some of the EMDS), it is possible that certain EMDS are queried by humans and that data are entered manually into the system

39. The system should 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.

40. Depending on the characteristics of the IMDS, it is highly advisable that the system is able to feed back to the participating institutions' database the information found during CoSAP and ADiSP. In case this cannot be done automatically (for technical reasons or for reasons of access privileges), the system will report all information back for manual (or supervised, i.e. non-automatic) ingestion.

41. When CoSAP and ADiSP are successfully completed the following basic answers shall be produced by the system:

     a. Is the Work in the Public Domain?
     b. Is the Work in-Copyright?
     c. If so, can the rights-owners be identified?
     d. If so, can all or some of them be located?
     e. If not, is therefore the Work Orphan or partially Orphan?
     f. If the Work is in-Copyright, the year when it will become Public Domain

42. The system will display the above listed results, as long as the detailed process of how these results were obtained, to the Internal User with privileges authorizing her/him to validate the results (*Validating User*) and, if the case, declare the Work Oprhan.

43. It is necessary to foresee that the results of the CoSAP and ADiSP are not conclusive, or that the Validating User does not consider the results clear enough. In this case, the Validating User shall be able to record the inability to come to a conclusive answer.

44. The system will have statistical functionalities designed in order to assess the efficiency of the system itself, but also available to Users for their internal reporting, or other purposes.

2. FUNCTIONAL REQUIREMENTS

2.1 Actors

2.1.1 Admin: Administrator user

The whole system will have an Administrator (Admin) who will manage and control the system in its integrity.

Admin rights are formally assigned by FORWARD's governance. Admin manages the system on behalf and under the instructions of FORWARD's governance.

For the duration of the project, the FORWARD governance corresponds to the project's partners according to D1.1 and the Consortium Agreement (where applicable).

Admin rights are managed by CINECA for the duration of the project.

Admin authenticate all users and user groups and assigns privileges to users, accounts and groups according to rules set by the project's governance.

The Admin function can obviously be separated by the technical maintenance of the system (although for the duration of the project these will obviously overlap in the partner CINECA).

2.1.2 BU-Beneficiary User: group

As described above, core user group of the FORWARD system are those institutions that according to the OWD and national transpositions are defined as 'beneficiaries' of the OWD.

Beneficiary Users (BU) have therefore the following characteristics:

- They are legal bodies and not individuals
- They are publicly accessible libraries, educational establishments and museums, archives, film or audio heritage institutions and public-service broadcasting organisations who:
  - o have a public mission
  - o are not-for-profit
  - o have in their collections works that are potentially Orphans (i.e. beneficiaries can only declare Orphan a work that is their collections)
  - o is formally listed as a beneficiary, if the law so requires in its MS,

      o   is authorized, as a consequence of the above, to declare a Work Orphan

Upon instructions of FORWARD governance, Admin will create, authenticate and, if required, eliminate a BU.

From a system point of view, it is necessary that the system can identify two types of BUs: Internal (IBU) and External (EBU).

IBU are defined as BUs who are members of the FORWARD entity (for the time being the project, later a legal entity to be defined). Membership to FORWARD implies, among other things, the federation of metadata resources. In this sense IBUs are expected to have a wider range of privileges than EBUs, although these restrictions will be better defined once the system, its governance and business models will be finalized.

Following instructions of the FORWARD governance, Admin will define a BU as Internal of External (needless to say, the status can change, in the two directions) and authorize/deny a given set of privileges.

A BU have a set of privileges and authorizations that include:

    a. Define, authenticate, maintain MS-Law, MS-DSR of its competence (can be shared with other BU of the same MS)
    b. Initiate Query – can submit, individually or in bulk, queries to the system
    c. Input data to Query - interacts with system (e.g. feeding requested info),
    d. Manage Queries – manages queries, can abort query(ies) (e.g. when information is insufficient)
    e. Status validation – validates copyright status, declares Work as Orphan
    f. Read process – has access to record of the whole process(es), cannot interact (e.g. to evaluate righsholders' claims)
    g. Read results – has access to results after process completed (validation), cannot interact

A BU cannot:

    a. Manage or in any way modify the overall FORWARD system (although certain tasks can be delegated by the Admin to BUs)
    b. Modify in any way MS-Law, MS-DSR that are from another MS
    c. Access queries or read processes initiated by other BU

IBUs (can) act as 'clearing centers' as defined in the DoW, in the sense that they can also perform queries on behalf of other beneficiaries, not (yet) registered in the system. These queries are handled in the same way as the ones originating from the IBU.

2.1.3 External users

The FORWARD system also foresee external users (*External User - XU*) that are not partners in the project or (in the future) affiliated to the FORWARD system.

These include other services and systems (e.g. ARROW) who are *not* beneficiaries, but who are given a certain degree of access to the system.

The OW database at OHIM will be the first instance of an XU.

Admin will manage XU accounts and groups (if applicable).

Level of privileges and access to the system for XU will be defined at a later stage.

It is also possible that these will be decided after the end of the project, and on a case-by-case scenario. For example, it is possible to imagine that FORWARD comes to an agreement for mutual access to an EMDS; another scenario is that FORWARD pushes certain metadata towards public databases like TEL or VIAF.

XU can also be defined as XU-groups, for example in the case of large, multi-institutions digitization projects. In that case, the project as such will be considered an XU, and the individual institutions participating in it will be authorized as accounts of the same group.

FORWARD will also maintain a publicly available list of Orphan works, similar to the one at OHIM. Users of this database will be able to search for a work and check its status. Only "positives" (i.e. works that were declared Orphans) will be included in this database. Users of these database are by definition EU.

2.2 Sources

The FORWARD system will manage the assessment of the CopyRights and Orphan status and the diligent search process of an AV work using some information that the system itself needs to receive from several sources. In D4.1 the required information was grouped in four categories:

- Information about film works
- Information about persons
- Information about corporations
- Information about right status of film works

Several resources maintain the above data and part of them are, from a system point of view, Internal Metadata Sources (IMDS), while others are External ones (EMDS).

2.2.1 Internal Metadata Sources

IMDS derives from resources owned and made available by the FORWARD partners. At present, these are limited to the resources of the 10 project's FHI partners, but they are expected to grow.

The deep analysis conducted in D4.1 higghlighted that each FHI may have information pertinent to the four categories mentioned above and the quality of the data in each archive was evaluated considering three aspects:

- Completeness
- Consistency
- Usability

All the necessary data enrichment process, required to improve the quality of their data and to be processed correctly, will be carried out by each IBU prior to making metadata available to the FORWARD system as foreseen in WP5. The Metadata Enrichment analysis, currently in progress has the aim to answer to some specific requirements such as:

Consistency: possible issues are that the information is not strong typed. In this case the FHI will try converting to the appropriate type or map their data into the correct type field of the FORWARD metadata schema.

Completeness: in principle, prior to the submission, each FHI will add the fields missing to the system of the clearing centre. There may be some cases that imply that some fields are enriched and maintained

in the FORWARD system. The choice between these two actions must be scoped on a field by field basis; some fields are only relevant to the FORWARD-system, whereas other fields are also relevant within the context of the applications of the clearing centres.

After the metadata enrichment, the FHI will be ready to make data available to the FORWARD system.

There are two different alternative ways for BUs to make data available:

1) through an initial load: each archive will feed the FORWARD system through a single or progress export of data present in their databases

2) through iterative submissions: a submission of a request performed by the BU constitutes a way to input structured data in the FORWARD system and can be achieved through a B2C interface.

During the design of D4.3 both alternatives will be evaluated further and the most appropriate strategy will be selected for the FORWARD system.

One important requirement is 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. In order to build an internal database of persons and organisations, it is necessary that each FHI exports the authority file for each data category. In this case the archive will provide initially the information related to the persons and the organisations present in their database, identifying each entity with a unique id used within the archive (if standards identifiers like ISAN or EIDR are not yet adopted). FHI will then provide the metadata related with the AV work; in the metadata, the persons and organisations associated with the filmographic record will be referenced using the respective internal ids.

In §5.2.2 of D4.1 the number of records for each data category of each FHI has been reported. The system will be designed in order to handle the whole catalogues. In any case, before the end of the system design phase, each FHI will define how many AV titles will be submitted to the FORWARD system in order to assess their Copyright and Orphan status. FHI will define whether they will export/submit just a subset of their complete catalogue (e.g. restricting to a time period, or a country of production,…) or their whole catalogue.

2.2.2 External Metadata Sources

External sources of information (EMDS) include all the other sources. These may belong to two different groups:

- **Sources defined by law**
- **Sources identified by FORWARD**

Sources defined by Law: According to the OWD, "the sources that are appropriate for each category of AV works or phonogram in question shall be determined by each MS, in consultation with right-holders and users".

At present, only Germany, Finland and The Netherlands have lists of sources determined by law. Other countries will provide them as soon as an implementation of the Orphan Works Directive will be completed.

It is also necessary to mention that the OWD, among the possible sources that each MS will define, indicates in the Annex a list of categories of sources that should be at least taken into consideration. They include:

(a) legal deposit;
(b) the producers' associations in the respective country;
(c) databases of film or audio heritage institutions and national libraries;
(d) databases with relevant standards and identifiers such as ISAN (International Standard Audiovisual Number) for audiovisual material, ISWC (International Standard Music Work Code) for musical works and ISRC (International Standard Recording Code) for phonograms;
(e) the databases of the relevant collecting societies, in particular for authors, performers, phonogram producers and audiovisual producers;
(f) credits and other information appearing on the work's packaging;
(g) databases of other relevant associations representing a specific category of rightholders.

Based on the applicable MS-specific workflow, the FORWARD system will configure the list of sources defined by law for that specific country (in the MS-DSR). Considering the high number of sources that might be necessary and the technical and/or procedural issues in accessing them (i.e. the system might not be authorized to automatically access some of the EMDS), it is possible that certain EMDS are queried by humans and that data are entered manually into the system. In these cases, for each data category, the system will just provide the link to the external resources that should be consulted in order to perform a diligent search.

During the course of the project it will be analysed if some of these sources can be integrated in the FORWARD system in order to be queried automatically by the system and, for these cases, the metadata structure and communication protocols to be adopted will be defined.

Sources defined by FORWARD - The legal questionnaires of D3.1 analysed which sources of information each FHI usually uses to perform its diligent search routine process. A cross answer check highlighted the common use of some international aggregation databases and services (such as VIAF). These results were integrated with further preliminary explorations and tests performed by CINECA on international open databases.

Those sources identified as relevant in the AV domain for the determination of the metadata associated to authors, corporations and films are actually under evaluation in order to ascertain whether they already have in place some B2B querying mechanisms that can facilitate integration within the FORWARD system.

Similarly to Arrow, the idea is to integrate in the FORWARD system these cross-border sources of data to automatize as much as possible the discovery of missing data during the FORWARD workflow.

The advantages to use international aggregators include:

- By enabling the FORWARD system to query a single access point to aggregated catalogues, the complexity of the FORWARD system can be reduced significantly as interoperability would only be necessary between the FORWARD system and each international aggregator.

- As these aggregated databases network extends to a large quantity of data sources spread around the world, we obtain the double advantage to scale the FORWARD system to all the countries in the European Union and to be able to perform cross border diligent searches required in case of co-productions, as stated in the previous chapters.

Example of international aggregations that are under analysis as EMDS include:

1. VIAF:

The VIAF (Virtual International Authority File – www.viaf.org/) combines multiple name authority files into a single OCLC-hosted name authority service. The goal of the service is to lower the cost and increase the utility of library authority files by matching and linking widely-used authority files and making that information available on the Web. Few years ago OCLC released VIAF in Linked data available under the Open Data Commons Attribution License (ODC-BY).

2. IDA:

The International Documentation on Audiovisual (IDA) works is a worldwide audiovisual rights management system and online database that Authors Societies consult to get accurate information on audiovisual works and rights owners.

Each work registration contains a minimum set of information: original titles, foreign titles, subtitles, production companies, year and country of production, shooting languages, rights owners, ISAN (International Standard Audiovisual Number), IPI (Interested Party Information), exploitation purposes and a unique IDA code.

CMOs use IDA to identify the transmitted audiovisual production and rights-owners, retrieve and treat the information before distributing the royalties to the identified sister organizations.

IDA is owned by CISAC (the International Confederation of Societies of Authors and Composers)

3. Wikipedia/DBPEDIA

The main knowledge archive contain a significant amount of information about persons. DBPEDIA is the Linked Open Data version of the free encyclopaedia and can be used in order to retrieve data automatically.

4. IMDB

One of the largest DB about movies and data related to cinema, IMDB it is often used as authority files even if its private nature make the use of the data not easy and expensive.

5. Freebase

Freebase is a community-curated database of well-known people, places, and things. It exposes as Open Data lots of information about films, people and organizations. Interlinks between Wikpedia and Freebase topics exists, making the use of this huge database easier.

2.3   *Main User Stories*

The aim of the present paragraph is to introduce the main user stories of the FORWARD system. A user story is a very slim and high-level definition of a requirement, containing just enough information so that the developers can produce a reasonable estimate of the effort to implement it.

User stories will be progressively updated all along the project according to the development of the Business and sustainability models that will be described in D2.3.2 and D2.3.3.

Some of these User Stories can be already described here and represent the user stories with high priority.

2.3.1 US1: Orphan Work status / IBU, EBU

Typically, this query will be initiated by a BU, being it Internal or External.

The query will be initiated via the input of filmographic data in the format that will be defined in D4.3. It is possible that queries come individually or in bulk (batch processing). Batches can be organized randomly, although well structured batches are preferable for efficiency reasons (e.g. works by a given author, or producer…).

The system will automatically interact with appropriate IMDS and EMDS, will require human intervention if necessary and will produce a set of results that will help the BU to declare the film Orphan (validation).

After validation, the system will follow a pre-set script to publish the information (e.g. directly to OHIM, to OHIM via NCA-National Competent Authority- X or Y, in parallel to OHMI *and* NCA, etc.).

2.3.2 US2: Orphan Work status / XU

FORWARD foresees that XU, non-BU, are interested in assessing the status of a work that is potentially Orphan.

Now, considering that the Orphan status assessment of a Work is a delicate matter, in terms of competence required, resources to be queried, and liability, it is projected that these requests will be initiated via one of the IBU that will act as providing an "expertise" as a service.

XU will naturally not be able to validate the Orphan status of a given work but are allowed to receive the results of the process.

2.3.3 US3: Copyright status query / IBU, EBU

IBUs and FORWARD members in general will be able to use the system also to assess whether a work is in-Copyright or not. As described above, any ADiSP will include a preliminary CoSAP, but the system will allow IBU to request also simple CoSAP.

Results will be fed back to the IBU who originated the query.

Depending on their relations with FORWARD (project now, entity later), and FORWARD's sustainability plans, EBUs will be able to do the same. It is very likely that this will be a for-a-fee service.

2.3.4 US4: Copyright status query / XU

As the copyright status is not dependent on the beneficiary status of the body or individual entering the request, this is likely to be the most sought-for among FORWARD services.

XU will be able to query the system; it is not decided yet whether this query can be submitted directly to

the system by the XU, or if it has to go through one of the IBU, acting as a "clearing centre". Again, complexity and liability issues are likely to indicate that the 'clearing centre solution" is probably to be preferred. Having said that, it is also possible that different models will be applied to different MS.

2.3.5 Other queries

It is to be foreseen that FORWARD will be also used to retrieve other information that are by-products of its basic functions.

These can include:

- information about authors including dates of birth/death
- information about representation of authors by which CMO
- information on producers and/or production companies

This information will be made available to XU either by the system or by the individual FORWARD members according to business rules that will be defined by the FORWARD governance.

In this context it is important to remind that not only the system, but also most of the resources (databases) on which the above information (as well as Orphan and in-Copyright status) is based, are owned by the FORWARD members.

Users of such information include the same institutions listed in the OWD but also the AV industry in general: broadcasters, production companies seeking to re-use existing AV works, individual researchers and users, CMOs, and so on.

3. NON FUNCTIONAL REQUIREMENTS

In addition to the obvious features and functions that the system will provide, described until now and known as functional requirements, there are other requirements that are not directly part of the FORWARD-specific core of functionalities and therefore are defined as "non-functional requirements" (NFR), although they are rather important in developing an efficient System.

The definition for a non-functional requirement is that it essentially specifies **how the system should behave. It can be also thought as a 'quality attribute' of the system.**

Non-functional requirements cover all the remaining requirements that are not covered by the functional requirements. They specify criteria by which the operation of a system can be evaluated, rather than its specific behaviors.

In the FORWARD system, it is not trivial to identify and detail all of the NFRs. This section tries to complete the description of non-behavioral requirements, sometimes implicitly already described in this document, with a short list of common areas to which they belong.

3.1 Performance

A good approach to deal with system performance issues is trying to answer to a few common questions.

What are the requirements about resources required, response time, transaction rates, throughput or anything else having to do with performance?

All required resources available to accomplish the goals of the project have to ensure an acceptable responsiveness also in case of peaks. The system expects a relatively low number of system users processing data simultaneously. Each heavy elaboration, such as bulk submissions, has to be limited with upper-bound thresholds and balanced properly to guarantee a good workload distribution and to avoid undesirable long waiting times.

Maximum data load is approx. 350 thousand titles to be processed for diligent search during the project period or immediately after the software release. This amount will be properly scheduled in order to balance the workload. Nevertheless, the system must provide a way to properly queue requests and distribute the load.

In certain cases, the system cannot ensure a short execution time because it could involve invocation of external services and even less because of manual intervention in the workflow.

Data load expected for authority files is approximately 200.000 records.

3.2 Security

The system must provide a security level in order to protect data and disallow undesirable intrusions. Users must be authenticated before accessing the system. Some functionalities will be restricted by default (e.g. users are allowed to manage only their own data) or allowed only under specific sets of

permissions (see above).

No other particular security requirements are foreseen; in particular no secure channel is required to protect transmitted data (i.e. no HTTPS).

3.3   Maintainability

The system needs to be **cost-effective to maintain**. These requirements address the costs of repairs as well as repair time. Maintainability is defined as the ease with which changes can be made to the FORWARD system. These changes may be necessary for the correction of faults or adapting of the system to meet a new requirement. The architecture, design, implementation, and documentation of the application shall minimize the maintenance costs.

Because maintainability requirements can be difficult to quantify, they are specified more as desired goals than as requirements that can be validated. Nevertheless, they could be specified in terms of architectural, design, and implementation constraints and the use of industry best practices that tend to produce maintainable applications and components when followed such as:

- Layered architectures.
- Modular software.
- Information hiding of implementation.
- Well-defined interfaces.
- Object-orientation and component-based development.
- Complete and current documentation.
- Adherence to project conventions
- Test plans and regression testing

3.4   Scalability

In general, scalability means the capability to cope and perform under an increased or expanding workload. A system that scales well will be able to maintain or even increase its level of performance or efficiency when tested by larger operational demands.

Is this really a requirement in Forward at the beginning? Probably no. The project expects to deal with approximately 500.000 items (including titles, persons and corporate) and, at present, no significant increase is foreseen. As well as, regarding the number of user having access to the system, it's not expected to significantly increase: the system will deal with a few administrators for each country.

4. CONCLUSIONS AND NEXT STEPS

This document concludes the activity of WP3.

It contains the blueprint of the FORWARD system as it outlines its basic requirements and functionalities. These requirements and functionalities as well as the actual FORWARD Workflow are further developed and described in D4.2.

The actual System set-up that is the objective of WP6 will become possible by the completion of the work leading to D4.3 "Design of rights information infrastructure".

The interaction of FORWARD with different types of XU depends on several factors and will be the object of further development of the Sustainability Plan.

Nevertheless, the blueprint here describes, covers the functionalities necessary to respond to all use cases that are likely to arise. Needless to say, the system must be defined in a realistic way so that it can be developed properly around its core functions. We are also confident that the system will be stable and flexible enough to allow development of further functionalities, if and when the need arises.

Although the actual workflow might need to be modified in some details when appropriate legislation on Orphan Works is implemented in all MS, we are also confident that functionalities and requirements contained in this document will hold in general terms.

GLOSSARY

ADiSP	Assisted Diligent Search Process
BU	Beneficiary User: a User of the System who is also a "Beneficiary" according to the definiton contained in the OWD
CMO	Collective Management Organizations
CoSAP	Copyright Status Assessment Process
EBU	External Beneficiary User: a BU who is not partner or member of the FORWARD System
EIDR	Entertainment Identifier Registry
EMDS	External Metadata Source: metadata resources external to the FORWARD System
FHI	Film Heritage Institution
IBU	Internal Beneficiary User: a BU who is also partner or member of the FORWARD System
IDA	International Documentation on Audiovisual
IMDS	Internal Metadata Source: metadata resources internal to the FORWARD System
IPI	Interested Party Information
ISAN	International Standard Audiovisual Number
MS	Member State
MS-DSR	Member State- Diligent Search Rules: Legislation and/or regulations about Diligent Search as defined in any given Member State
MS-Law	FORWARD-relevant legislation on Copyright in any given Member State
OHIM	Office for the Harmonization of the Internal Market
OW	Orphan Work
OWD	Orphan Works Directive
PD	Public Domain
TEL	The European Library
VIAF	Virtual International Authority File
XU	External User - a User who is extrenal to the FORWARD System and is not a Beneficiary