MIRROR
Migration-Related Risks caused by misconceptions of Opportunities and Requirements

Grant Agreement No. GA832921

Deliverable D7.2

<table>
<thead>
<tr>
<th>Work-package</th>
<th>WP7: Architecture and Information Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliverable</td>
<td>D7.2: MIRROR Information Model: Initial Version</td>
</tr>
<tr>
<td>Deliverable Leader</td>
<td>EURIX</td>
</tr>
<tr>
<td>Quality Assessor</td>
<td>FOI</td>
</tr>
<tr>
<td>Dissemination level</td>
<td>Public</td>
</tr>
<tr>
<td>Delivery date in Annex I</td>
<td>M12, May 31, 2020</td>
</tr>
<tr>
<td>Actual delivery date</td>
<td>May 31, 2020</td>
</tr>
<tr>
<td>Revisions</td>
<td>v1.0 Document revised after QA comments</td>
</tr>
<tr>
<td>Status</td>
<td>Final</td>
</tr>
<tr>
<td>Keywords</td>
<td>Conceptual Information Model, Media Element, Annotation, Finding, Observed Context, Observed Situation, Threat, Perception</td>
</tr>
</tbody>
</table>
Disclaimer

This document contains material, which is under copyright of individual or several MIRROR consortium parties, and no copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the MIRROR consortium as a whole, nor individual parties of the MIRROR consortium warrant that the information contained in this document is suitable for use, nor that the use of the information is free from risk, and accepts no liability for loss or damage suffered by any person using this information.

This document reflects only the authors' view. The European Community is not liable for any use that may be made of the information contained herein.

© 2020 Participants in the MIRROR Project
### List of Authors

<table>
<thead>
<tr>
<th>Partner Acronym</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>EURIX</td>
<td>Francesco Gallo, Germano Russo</td>
</tr>
<tr>
<td>LUH</td>
<td>Erick Elejalde, Claudia Niederee, Miroslav Shaltev</td>
</tr>
<tr>
<td>SAIL</td>
<td>Gerhard Backfried, Erinç Dikici</td>
</tr>
<tr>
<td>CERTH</td>
<td>Vasileios Mezaris, Alexandros Pournaras</td>
</tr>
</tbody>
</table>
**Table of Contents**

Executive Summary ........................................... 5

1 Introduction .................................................. 6
   1.1 Purpose of the MIRROR Information Model .......... 6
   1.2 Target Audience ........................................... 7
   1.3 Structure of the Deliverable .............................. 7

2 Conceptual Information Model .............................. 8
   2.1 Foundations of the Conceptual Model .................. 8
   2.2 Components of the Conceptual Information model .... 9
       2.2.1 Media Elements ........................................ 9
       2.2.2 Annotations ............................................ 11
       2.2.3 Findings and Observed Contexts/Situations ......... 13
       2.2.4 Perception and Threats ............................... 15
   2.3 Complete Model Overview ................................ 16
   2.4 Core Processes ............................................ 17

3 Information Model in the Framework Prototype ............ 20

4 Conclusion .................................................... 22

5 Annex: List of all Elements of the Conceptual Information Model ........................................... 23

6 References ...................................................... 26

7 Acronyms ......................................................... 26
Executive Summary

The main objective of this deliverable is the definition of the initial version of the MIRROR information model, referenced to in the following as the MIRROR Conceptual Information Model (MIRROR-CIM), whose main purpose is to capture the project’s understanding of the core concepts about the analysis of the migration-related perceptions and misperceptions of Europe by migration actors and the resulting threats. Such perception is driven by the consumed digital and non-digital representation of partially biased information related (in the widest sense) to the migration context.

The definition of the MIRROR-CIM improves the shared understanding of the concepts within the project interdisciplinary team and clarifies the conceptual understanding (and terminology) of the project for external readers as a basis for better understanding other parts of the project, making use of relationship diagrams and definitions for all model elements.

The core entities of the model are Media Elements, Annotations, Perception-related Findings (e.g., Detected Events and Information Campaigns), Observed Contexts and Situations, Perception and Threats. In addition to this set of Core Elements, the MIRROR-CIM contains also a number of other relevant elements, which are linked to the Core Elements and refine and contextualize them.

The MIRROR-CIM connects the basic entities associated to collected data, the Media Elements, with higher level concepts related to the End User observation, such as the Contexts and Situations. The MIRROR system collects and analyses Media Elements, the digital representations of partially biased migration related information. Different types of media analyses are applied, generating the Annotations, which provide the link of the Media Elements to further information items. In addition to this, Media Elements are further processed using summarization techniques to ease digestion (the result is represented in the model by a Summary entity). Based on further processing and aggregation of the results of the media analysis mentioned above, higher level entities are introduced in the model, such as the perception related Findings. The types of Findings which are considered in the MIRROR-CIM are tailored to the goals and scope of the MIRROR project, namely the analysis of Perception and the identification of possible Threats. In the MIRROR approach, such Threats are triggered by migration-related perceptions and especially misperceptions. Moving to the highest level in the model, two types of entities are defined, namely the Observed Contexts and the Observed Situations. The End User perceives the MIRROR results in terms of such Observed Contexts/Situations, which combine and select Media Elements, Summaries, Annotations and Findings according to the users’ preferences.

Based on the current understanding in the project, a number of core processes have been identified as associated to the entities of the MIRROR-CIM, namely Data Collection, Pseudonymization, Media Analysis, Observed Context Selection, Cross-media Analysis (which includes Summarization), Deletion, Verification and Threat Linkage. For each of them a definition and a process diagram are provided.

As already mentioned, the main purpose of the initial version of the model was not a comprehensive and detailed definition of all entities, relationships and processes, but rather the need for framing the different ideas in the project and support the awareness of the big picture, targeting the model final version (M27).

The architecture defined in D7.1 and the MIRROR-CIM described in this document provide input to the implementation of the first MIRROR prototype (D7.3). In particular, it is worth mentioning that the model strongly influenced the design of the framework User Interface (UI), which provides the End User not only with information about high level concepts such as Observed Contexts/Situations and Findings, but also with the details about model core elements, such as the Media Elements, Annotations and Summaries.
1 Introduction

The current deliverable provides a description of the initial version of the MIRROR Information Model, based on the current understanding of the project in terms of main scenarios, requirements, and project outcomes, in particular the MIRROR Framework. The main purpose of this document is to set the foundations: a common vocabulary of core model elements and the first design of the main relationships and processes involving such elements. The core model elements are used to represent the acquisition of data relevant to the analysis of the EU perception in the Country Of Origin (COO) of migrants, the information extracted from the different types of media content and the results of the cross-media, in order to relate such information to analysis of perception and threats.

Following the initial version of the MIRROR Information Model, in parallel with the MIRROR prototype implementation, the collection of feedback from the users and the progress in all project WPs, the final version of the model will be defined and included in D7.4, at project month 27.

The main challenges in the definition of such a model are due to the complexity of the phenomena to be analyzed and to the variety of technologies which will be developed. In particular, the model is expected to: (1) support the representation of the different data sources, (2) express the combination of the results from the automated media analysis and from empirical studies and (3) drive the development of the framework.

The input to this deliverable is provided by D2.1 and in general from WP2, in terms of scenarios and requirements, and also by the architecture defined in D7.1.

The approach adopted in the definition of the model is to first identify a set of core entities and their fundamental relationships, but with enough flexibility to extend such core elements in order to support the specific models associated to the different types of analysis carried out in WP4, WP5 and WP6. In addition to the elements and their relationships, the model aims also at defining the relevant workflows associated to data exchange, transformation and persistence.

The information model described in this deliverable is being used as input for the design and implementation of the framework, in particular for the data layer and the interfaces exposed by the MIRROR framework. Hence, the model will be used to define both the representation of data persisted within the MIRROR framework and the data exchanged with the integrated tools and applications.

Concerning the representation of the model, different standard modelling languages could be used, such as UML2 [OMG, 2017] (model entities) and BPMN [OMG, 2011] (workflows). For the sake of simplicity and also considering that the model will evolve towards the final version throughout the whole duration of the project, in this deliverable the BPMN representation has been used for the processes, whilst for the core entities and their relationships no particular standard representation has been used.

1.1 Purpose of the MIRROR Information Model

The main purpose of the Information Model is twofold: (1) organize in a coherent representation the core elements required to represent all the relevant data and the extracted information retrieved, analyzed and stored in the MIRROR framework and then provided to the user; (2) define the main processes involving data and their transformations to fulfill different requirements. The model includes the main outcomes of the different WPs and is intended also as a proof of concept of the MIRROR approach. Moreover, the framework development is driven by the model: the technical components, the scenarios and the integration framework will support the core elements and the processes defined in the model.
The initial version of the model described here is mainly focused on core elements and on the current understanding within the project. The final version of the model is expected at project month 27 and will be built on top of the foundations described here.

1.2 Target Audience

Since the main purpose of this model is to gather information and awareness from the project in a coherent framework, the content of this deliverable is neither technical nor tied to any specific implementation, hence the audience of the document is expected to be quite broad.

1.3 Structure of the Deliverable

The deliverable is organized as follows: the core part of the model, i.e., the MIRROR-CIM, is described in Section 2; the relationship with the framework prototype and other components is described in Section 3; the Conclusion is in Section 4; a Glossary with the definitions of the model elements is provided in Section 5.
2 Conceptual Information Model

The scope of this deliverable is to define the information part of the overall MIRROR model. In the context of the project, this part is identified by the term MIRROR-CIM. The MIRROR-CIM captures the project’s understanding of the core concepts relevant for the MIRROR project. The development of the MIRROR-CIM is thus both a practice in a) improving the shared understanding of the concepts within the MIRROR-project interdisciplinary team and b) clarifying the conceptual understanding (and terminology) of the project for external readers as a basis for better understanding other parts of the project.

In Section 2.1 we describe the foundations of the MIRROR-CIM and the derived Core Elements. In Section 2.2 the components of the MIRROR-CIM are discussed in more detail. Section 2.3 presents the full MIRROR-CIM with all elements and relationships defined. Complementing the structural part, Section 2.4 presents the most important media-related workflows/processes in the MIRROR project.

Although the MIRROR-CIM has been thoroughly discussed in the consortium, it still has to be considered as a work in progress: the conceptualization of the relevant domain within the project will still evolve further based on the experiences made in the project.

2.1 Foundations of the Conceptual Model

For keeping the model MIRROR-CIM concise and closely linked to the MIRROR project goals, it was decided to base it on the key mechanisms, which the project MIRROR wants to analyze.

The left part of Figure 1 summarizes those key mechanisms: migration-related perceptions and misperceptions of Europe by migration actors and the resulting threats are the main subjects of the analysis.

Figure 1: Representation of the migrants’ perception mechanisms analyzed in MIRROR (left side, on white background) mapped to the Core Elements of the MIRROR-CIM (right side, on light-blue background)
Such perception is driven by the consumed digital and non-digital representation of partially biased information related (in the widest sense) to a migration context. In addition, perception is also strongly influenced by the societal and political context to which a migration actor belongs, which has a strong impact on the framing of the consumed information. Perception and misperception influence migration decisions and might lead to migration related threats (again also influenced by the societal and political context). The target of the MIRROR project is to enable end-users of the planned MIRROR system to estimate and understand the evolving perceptions and misperceptions, the threats that might result from them and the possible counteractions.

The key mechanisms described above have been used to derive the Core Elements of the MIRROR-CIM. In Figure 1 this process of derivation is depicted, showing on the right side the derived Core Elements with the main mechanisms connecting them.

In a nutshell, MIRROR collects and analyzes Media Elements (digital representations of partially-biased, migration-related information). The application of different types of media analysis methods leads to Annotations, which link Media Elements to further information items. In addition Media Elements are summarized to ease digestion (Summary). Further processing and aggregation steps lead to higher-level perception-related Findings. The types of findings considered (e.g., Information Campaigns) are tailored to the goals of the MIRROR project, namely the analysis of Perception and the identification of possible Threats triggered by migration-related perceptions and especially misperceptions. The End Users perceive the MIRROR results in terms of Observed Contexts/Situations, which combine and select Media Elements, Annotations, Summaries and Findings according to End Users’ preferences.

In addition to the set of Core Elements shown above, the MIRROR-CIM contains a number of other relevant elements, which are linked to the Core Elements and refine and contextualize them.

### 2.2 Components of the Conceptual Information model

The structure of the Information model sets the elements of the MIRROR-CIM in relationship to each other. The following types of relationships are distinguished: hierarchical relationships (is-a), composition relationships (has-a) and general relationships. The notation used for their representation is derived from Unified Modeling Language (UML) [OMG, 2017].

For easing the digestion of the MIRROR-CIM, individual areas of the model are discussed separately starting from the overview of the Core Elements presented in Section 2.1. In the rest of this section the individual parts of the MIRROR-CIM are discussed.

#### 2.2.1 Media Elements

The Media Element is one of the Core Elements in perception analysis and thus also in the MIRROR-CIM. Figure 2 shows the part of the MIRROR-CIM revolving around the Media Element concept.

The elements contained in this part of the MIRROR-CIM are defined as follows:

**Media Element** An information object coming from media, which is considered as one unit by the processing applied to it. A Media Element can be a Media Item, a Mixed Media Item or a Media Collection. The introduction of this more general object eases the overall modelling. A Media Element is the object to which Annotations are assigned in the project.
Media Item  An object containing a unit of information of one Media Type such as an Audio, an Image, a Text or (a part of) a Video.

Mixed Media Item  An information object composed from several other Media Elements, e.g., a tweet containing image and text or a video composed from a number of fragments.

Media Collection  Collection of (typically uniform) Media Elements. A Media Collection is itself also a Media Element and thus, e.g., Annotations might be associated to it.

SM SubNetwork  A subnetwork extracted from a social media network. The SM SubNetwork is composed from Actors in the network, Media Elements and Relationships, representing links between them.

Media Type  The Media Item type. Currently the model considers the following Media Types: text, video, audio and image.

Relationship  Representation of a relation between objects of the domain. This can for example be Media Elements or Actors.

Language  The language(s) of a Media Element. This might refer to Text as well as to Audio and Video.

Origin  The country the Media Element is originating from.

Publication Time  The time point at which the Media Element was published/created. Alternatively, the retrieval time if the creation/publication time is not known.

Media Source  Source from which the respective Media Element was collected (such as CNN, DW, Twitter, Facebook...). The Media Source has among its attributes the Source Type, the Channel and the actual publisher (Actor); For social media content the actual contributors would be modelled as publishers, whereas the Channel captures the respective social media channel.

Channel  Example channels are a newspaper, a social media channel or a TV channel.

Actor  An Actor can be a Person, an Organization or an Account.

Source Type  The class or type of a Media Source. In the model we currently foresee the classes News and Social Media.
The MIRROR project analyzes media content coming from social media as well as traditional media. On an abstract level, an analysed media content is represented by a Media Element. The introduction of this more abstract concept eases uniform and concise modelling of the relationships. A Media Element generalizes four types of information objects (concept hierarchy): Media Items (as the simplest form), Mixed Media Items (as composites from several Media Elements), Media Collections (collections of Media Elements) and SM SubNetworks.

A Mixed Media Item, a Media Collection as well as an SM SubNetwork can contain other objects of type Media Elements - as part of their composition, as elements of the collection, and as part of the network, respectively (composition relationship). As part of its composition, an SM SubNetwork is additionally linked to Actors and Relationships connecting Actors and Media Elements within the sub-network. A Media Item is linked to a Media Type, which describes its type in terms of media (e.g., image).

A Media Element is linked to a number of properties. Within the model the properties Publication Time, Language and Origin (in terms of Location) are highlighted due to their importance for media processing and interpretation. Within the project Media Elements are coming from different sources such as Newspapers, TV-Channel or Social Media channels. Knowing the source of a Media Element can also contribute to its interpretation.

In the MIRROR-CIM the source of Media Elements is modeled by the element Media Source, which is associated with a Media Element. Each Media Source is associated with a Source Type, which distinguishes between traditional media (News) and Social Media. Furthermore, it is linked to an Actor (acting as its publisher) and a Channel (on which it was published). Note that the distinction between the publishing Actor and the Channel is very important in the context of Social Media, given its user-generated content.

### 2.2.2 Annotations

A core activity in the MIRROR project is analyzing Media Elements for extracting further information from them. This results in different types of Annotations. Figure 3 shows the part of the MIRROR-CIM revolving around the Annotation concept.

![Figure 3: Representation of the MIRROR-CIM part related to the Annotation concept](image-url)
The elements contained in this part of the MIRROR-CIM are defined as follows:

**Annotation**  An additional piece of information associated with a Media Element, typically derived by processing the Media Element, e.g., using methods of image or text analysis (possibly also taking into consideration further sources of information).

**Annotation Type**  The type of the Annotation, which describes the origin of the Annotation. The Annotation can be: a) extracted/derived (Derived); b) given as metadata by the provider (Metadata). A Derived Annotation could be for example a Person Entity extracted from Text, while an example of Metadata Annotation is the username associated with a social media post. Mixed cases also exist, e.g., when the provider has already used extraction methods to annotate the Media Element.

**Media Element**  An information object coming from media, which is considered as one unit by the processing applied to it. A Media Element can be a Media Item, a Mixed Media Item or a Media Collection. The introduction of this more general object eases the overall modelling. A Media Element is the object to which Annotations are assigned in the project.

**Entity**  The representation of a real world entity such as a Person or an Organization. An Annotation can link a Media Element to one or more entities.

**Person**  The representation of a real world person or a system user. Following the ethical principles of the project, working with real names will be avoided wherever possible (i.e., via application of the Pseudonomization process, treated in Section 2.4).

**Organization**  The representation of a real world organization.

**Actor**  An Actor can be a Person, an Organization or an Account.

**Account**  User account or user name, e.g., in a Social Media channel. Link to the actual person might not be known. There might be several user names for one person.

**Sentiment**  A Sentiment Annotation can be used to express the prevailing sentiment polarity and stance of a Media Element content.

**Topic/Concept**  Part of the conceptualization of the considered domain. An Annotation can associate a Media Element with one or more Concepts. An important type of Concept considered in the MIRROR project is the Migration-Related Semantic Concept (MRSC).

**MRSC**  Migration-Related Semantic Concept, that is, a Concept of specific interest in the migration context, such as “racism”, “opinion former”, “refugee”, “domestic service”, “education” (see the analysis provided in Section 5 of D2.1).

**Generic Concept**  Concept used in describing the content of a Media Element. Not directly related to migration.

**Caption**  Text linked to an Image or another Media Element describing the content of the Media Element.

**Timeframe**  A piece of information describing temporal information. An Annotation can associate a Media Element with a Timeframe. This time frame refers to the content of the Media Element (the time the content directly or indirectly refers to).

**Location**  Location refers to a place. Different types of granularity are possible here. An Annotation can associate a Media Element with one or more Location.
**Authenticity** Information that a Media Element is authentic. This can for example be related to manipulation of content and to artificial generation of content via bots (artificial origin of Media Element). An Annotation can associate a Media Element with information about its Authenticity.

**Relationship** Representation of a relation between objects of the domain. This can for example be Media Elements or Actors.

**Summary** A Summary is a piece of information derived from a Media Element (especially from a Media Collection) by summarizing it, typically using automated methods.

The core element Annotation relates a Media Element to other types of information. Within the model, the main relationship of an Annotation is thus the relationship with a Media Element. Furthermore, an Annotation is linked to one or more other information elements such as Entities, Locations, a Timeframe or Topics/Concepts. An Annotation can also be a Relationship or a set of Relationships, which have been extracted from one or more Media Elements. In addition, an Annotation can also link a Media Element to a Sentiment capturing the results of Sentiment Analysis. Note, that the way of modelling is quite flexible here. An Annotation for example also can link a Media Element, an Entity and a Sentiment for expressing the sentiment about this Entity found in the Media Element. Finally, an Annotation can also link a Media Element to information about its Authenticity.

### 2.2.3 Findings and Observed Contexts/Situations

A strong user orientation is core for the MIRROR system. Therefore, Observed Context/Situation, which compiles and updates relevant information for the End User, and Findings, which aggregate evidences found in media into higher-level insights, are also part of the Core Elements of MIRROR-CIM.

Figure 4 shows the part of the MIRROR-CIM revolving around the Finding and the Observed Context/Situation concepts.

![Figure 4: Representation of the MIRROR-CIM part related to the Finding and Observed Context/Situation concepts](image-url)
The elements contained in this part of the MIRROR-CIM are defined as follows:

**End User**  User of the MIRROR system. He/she uses the system to get informed about media activities in the context of the situation(s) he/she observes, the resulting expected Perceptions of this situation by migrants and the possible Threats that may result.

**Observed Context/Situation**  Part of the media activity and the findings generated by MIRROR analysis, which is relevant for the End User and is, therefore, observed. Observed Contexts/Situations are defined by the End User via filters and are filled by the system with relevant information; these are (annotated) Media Elements, Summaries and Findings related to the context.

**Finding**  Perception-related finding, which is created by combining evidences from annotated Media Elements. Examples of such Findings currently considered are Detected Event and Information Campaign. The types of Findings considered will still evolve.

**Information Campaign**  Campaigns via media, which provide information and also misinformation with the aim to influence migration-related decisions.

**Social Network Pattern**  Perception-relevant pattern learned from the analysis of SM SubNetworks coming from one or more Social Media (SM) Channels.

**Influential Actor**  Actor which is influential within SM SubNetworks of a SM Channel. In the project we are interested in Influential Actors in the context of migration decisions, Information Campaigns, Perceptions in the migration context and migration-related Threats.

**Detected Event**  Evolving or past event detected from evidences found in media Channels.

**Media Element**  An information object coming from media, which is considered as one unit by the processing applied to it. A Media Element can be a Media Item, a Mixed Media Item or a Media Collection. The introduction of this more general object eases the overall modelling. A Media Element is the object to which Annotations are assigned in the project.

**Annotation**  An additional piece of information associated with a Media Element, typically derived by processing the Media Element, e.g., using methods of image or text analysis (possibly also taking into consideration further sources of information).

**Summary**  A Summary is a piece of information derived from a Media Element (especially from a Media Collection) by summarizing it, typically using automated methods.

The End User observes and analyzes the media situation via Observed Context/Situation, which he/she has defined via filters. The End User thus has a twofold relationship to the element Observed Context/Situation. The End User defines the Observed Context/Situation (maybe also dynamically defining it) and He/she observes and inspects the content collected into the context by the system over time. Both activities might be performed by different End User roles in the system.

Three types of information are collected into Observed Contexts/Situations: Relevant Media Elements, Summaries created from (annotated) Media Elements and Findings.

Findings are obtained by combining and further processing the Annotations extracted from Media Elements. They are, thus, linked to Annotations and indirectly (via Annotations) to Media Elements. The extensible set of Findings considered in the project currently includes Detected Events, Information Campaigns, Influential Actors and Social Network Patterns.
2.2.4 Perception and Threats

Better understanding the perception and misperception of Europe in the context of migration and the threats that may result from these (mis-)perceptions is one of the main goals of the MIRROR project. The MIRROR-CIM elements associated to the Perception and Threat Core Elements are shown in Figure 5 and are defined as follows:

**Perception**  Perception of Europe, individual European countries and European political and societal issues in the context of migration.

**Threat**  Migration-related threat, caused by the perception and misperception of migration contexts.

**Political & Societal Context**  Context in the environment of the respective Actor, which influences Perception, migration decision and migration-related actions.

**Observed Context/Situation**  Part of the media activity and the findings generated by MIRROR analysis, which is relevant for the End User and is, therefore, observed. Observed Contexts/Situations are defined by the End User via filters and are filled by the system with relevant information; these are (annotated) Media Elements, Summaries and Findings related to the context.

**Finding**  Perception-related finding, which is created by combining evidences from annotated Media Elements. Examples of such Findings currently considered are Detected Event and Information Campaign. The types of Findings considered will still evolve.

**Media Element**  An information object coming from media, which is considered as one unit by the processing applied to it. A Media Element can be a Media Item, a Mixed Media Item or a Media Collection. The introduction of this more general object eases the overall modelling. A Media Element is the object to which Annotations are assigned in the project.

**Annotation**  An additional piece of information associated with a Media Element, typically derived by processing the Media Element, e.g., using methods of image or text analysis (possibly also taking into consideration further sources of information).

**Actor**  An Actor can be a Person, an Organization or an Account.

**Account**  User account or user name, e.g., in a Social Media channel. Link to the actual person might not be known. There might be several user names for one person.

**Person**  The representation of a real world person or a system user. Following the ethical principles of the project, working with real names will be avoided wherever possible (i.e., via application of the Pseudonomization process, treated in Section 2.4).

**Organization**  The representation of a real world organization.

The main relationships are between Finding, Perception and Threat. Findings extracted from the analysed Media Elements (from traditional and Social Media) point to Perception of migration-related contexts and to potential Threats caused by such (mis-)Perception.

In analyzing migration-related Perceptions and Threats it has to be kept in mind that they are not triggered by the media consumed by Actors alone. Both also depend on the societal and political context, which contributes to the framing of the information consumed. Therefore, Political & Societal Context also has been included as an element into MIRROR-CIM. Furthermore, Perception and Threats are typically also linked to (groups of) Actors, which are associated with the Perception and the potential Threats, respectively.
2.3 Complete Model Overview

An overview of the whole MIRROR-CIM is depicted in Figure 6, including the parts of the model described in the previous section. The main purpose of this representation is to share in a compact form the whole model, to be discussed and refined in the project, while separate parts can be further detailed and refined.

Figure 5: Representation of the MIRROR-CIM part related to the Perception and Threat concepts

Figure 6: Complete representation of the MIRROR-CIM
2.4 Core Processes

In this section we briefly describe the core processes which have been identified so far, mainly concerning data management. Such processes are related to Data Management and Data Analysis layers of the MIRROR architecture, as discussed in Section 3.

It is worth mentioning that the list of processes mentioned in the following is not comprehensive and reflects only the current understanding. The current processes will be refined during the project, when new ones will be also added.

The core processes identified so far are:

- **Data Collection** Process responsible for the retrieval of Media Elements from the specified Media Sources.
- **Pseudonymization** Process by which personally identifiable information fields in Media Elements and Annotations are replaced by one or more artificial identifiers, or pseudonyms. The resulting mapping of pseudonyms to original names is kept in a separate secure repository (vault).
- **Media Analysis** Process that enriches with Annotations the Media Elements, performing the analysis for the different Media Types.
- **Observed Context Selection** Process by which an End User defines an Observed Context/Situations of interest.
- **Cross-media Analysis** Process that combines and further analyzes the Annotations extracted from Media Elements to produce Findings. It includes the **Summarization**, a process that creates a digested version of the analyzed content to ease access.
- **Deletion** Process that implements the data retention policies (data-access minimization to protect security and privacy) as set by the Administrators.
- **Verification** Process by which expert End Users can assess the validity of the analysis and provide feedback to the system.
- **Threat Linkage** Process that links the results of Media and Cross-media Analysis processes to potential Threats.

The processes above can be either implemented by Framework components or performed by a human.

For the representation of the processes we make use of the Business Process Model and Notation (BPMN), which is a standard notation proposed and maintained by the Object Management Group (OMG) to represent business processes [OMG, 2011]. The main advantage of such notation is that the diagrams are represented using a standard textual format, which enables interoperability among different tools: the rendering of the diagrams is then supported by each BPMN compliant tool, based on the same notation, but allowing additional features and customization for example in commercial tools.

In Figure 7 the main data management workflow is depicted. The lanes in the diagram represent the different actors (both humans and machines), which are responsible for the execution of one or more steps (boxes) in the process. Some steps have additional icons, emphasizing that the step is executed by a human or by a software component. The lines represent exchange of messages.
As shown in Figure 7, the admin user is responsible for configuring the MIRROR system for what concerns the setup of the media sources, which are then used to retrieve data relevant for the project analysis within the MIRROR framework; after the analysis process is completed, the MIRROR system user is notified.

Figure 7: Main data management workflow diagram in BPMN notation, involving three main actors: the MIRROR framework, a user with admin privileges and a regular user of the MIRROR system.

In Figure 8 an example is provided with a detailed description of one of the core processes mentioned above, the Data Collection. Using the same notation, the specific steps related to the collection and storage of data are represented.

Further diagrams associated to the other core processes will be generated and shared within the project and included in the final version of the model.
Figure 8: Main data management workflow diagram in BPMN notation, involving three main actors: the MIRROR framework, a user with admin privileges and a regular user of the MIRROR system.
3 Information Model in the Framework Prototype

The MIRROR-CIM described in the previous section is related to the MIRROR Framework architecture described in deliverable D7.1 and provides input to the implementation of the prototype, which represents a proof-of-concept for the assessment of the core concepts and results developed in the project.

An overview of the MIRROR architecture is depicted in Figure 9, based on the diagrams included in D7.1. The different layers of the architecture are represented.

The Data Management layer includes both the components responsible for data collection and storage, but also those in charge for implementing privacy and security processes, such as pseudonymisation and retention policies. Such components are activated according to some of the processes defined in Section 2.4, in particular Data Collection, Pseudonymisation, Deletion and Verification. The collected data are represented in the MIRROR-CIM by the Media Element concept as defined in Section 2.2.1.

The Data Analysis layer includes the components responsible for the different types of data analysis, supporting the different source types. The associated processes from Section 2.4 are Media Analysis, Linking and Summarization. The results of the analysis are mapped to Annotations, Links, different types of Concepts and other related model elements in Section 2.2.2.

The Access and Integration layers enable the transition from the data level to the user level, the Client Applications layer.

The role of the Access Layer consists in translating the technical information available in the framework backend into data accessible and understandable by the user, through high level APIs. In addition to this,
the Access layer takes into account the user role and profile and monitors the user activity: this information is useful to provide the user with the correct information and at the right level of detail, where privacy and security are guaranteed.

The Integration layer is in charge for the communication among the layers and of orchestration of the processes, hence is supposed to be running behind the scenes during the execution of the processes in Section 2.4.

The Client Applications layer, which communicates with the Data Analysis layer through the Access layer, is the user level: on top of the analysis results, adding a higher level of aggregation, the user can access the information provided by the framework. At this level the model elements which are relevant are the Findings, the Observed Contexts and Situations and all the other elements described in Section 2.2.3. Based on them, the analysis of the perception and misperception applies: this is represented in the model by Perception and Threats elements in Section 2.2.4
4 Conclusion

In this deliverable the initial version of the MIRROR Information Model has been defined, based on the core MIRROR idea that perception and misperception of Europe by migration actors and the resulting threats are driven by the consumed digital and non-digital representation of partially biased information.

The model includes core entities at the digital content level, namely the Media Elements, the Annotations created by different types of analysis and the Summary; at a higher level of aggregation the model introduces perception-related Findings, while at the End User level Observed Context/Situations become relevant, combining the lower elements according to the users preferences.

The model includes static diagrams representing the core elements and their relationships, as well as dynamic diagrams where the core processes are depicted. For each element in the model, a definition has been provided.

The next steps will be:

1. to validate the model against the first prototype implementation (project month 14),
2. to evaluate further requirements related to project scenarios (WP2) and the outcomes of WP3 from a legal and ethical perspective, to be integrated in the conceptual model entities and processes,
3. to improve the conceptual model based on the development of the technical tools from WP4, WP5 and WP6,
4. to integrate all contributions into the final version of the model (project month 27).

Being the initial version of the model, rather than aiming at a comprehensive description of all the elements and processes which will be investigated by MIRROR, the current definition of the MIRROR-CIM improves the shared understanding of the concepts within the interdisciplinary team and clarifies the conceptual understanding (and terminology) of the project for external readers as a basis for better understanding other parts of the project, targeting the final version of the model at M27.
5 Annex: List of all Elements of the Conceptual Information Model

Account User account or user name, e.g., in a Social Media channel. Link to the actual person might not be known. There might be several user names for one person.. 12, 15

Actor An Actor can be a Person, an Organization or an Account.. 10, 11, 12, 13, 14, 15, 23, 24, 25

Annotation An additional piece of information associated with a Media Element, typically derived by processing the Media Element, e.g., using methods of image or text analysis (possibly also taking into consideration further sources of information).. 5, 9, 10, 11, 12, 13, 14, 15, 17, 20, 22, 23, 24, 25

Annotation Type The type of the Annotation, which describes the origin of the Annotation. The Annotation can be: a) extracted/derived (Derived); b) given as metadata by the provider (Metadata). A Derived Annotation could be for example a Person Entity extracted from Text, while an example of Metadata Annotation is the username associated with a social media post. Mixed cases also exist, e.g., when the provider has already used extraction methods to annotate the Media Element.. 12

Authenticity Information that a Media Element is authentic. This can for example be related to manipulation of content and to artificial generation of content via bots (artificial origin of Media Element). An Annotation can associate a Media Element with information about its Authenticity.. 13

Caption Text linked to an Image or another Media Element describing the content of the Media Element.. 12

Channel Example channels are a newspaper, a social media channel or a TV channel.. 10, 11, 14, 23, 24, 25

Detected Event Evolving or past event detected from evidences found in media Channels.. 14, 15, 23

End User User of the MIRROR system. He/she uses the system to get informed about media activities in the context of the situation(s) he/she observes, the resulting expected Perceptions of this situation by migrants and the possible Threats that may result.. 5, 9, 13, 14, 15, 17, 22, 24

Entity The representation of a real world entity such as a Person or an Organization. An Annotation can link a Media Element to one or more entities.. 12, 13, 23

Finding Perception-related finding, which is created by combining evidences from annotated Media Elements. Examples of such Findings currently considered are Detected Event and Information Campaign. The types of Findings considered will still evolve.. 5, 9, 13, 14, 15, 17, 21, 22, 24

Generic Concept Concept used in describing the content of a Media Element. Not directly related to migration.. 12

Influential Actor Actor which is influential within SM SubNetworks of a SM Channel. In the project we are interested in Influential Actors in the context of migration decisions, Information Campaigns, Perceptions in the migration context and migration-related Threats.. 14

Information Campaign Campaigns via media, which provide information and also misinformation with the aim to influence migration-related decisions.. 9, 14, 15, 23
Language  The language(s) of a Media Element. This might refer to Text as well as to Audio and Video.. 10, 11

Location  Location refers to a place. Different types of granularity are possible here. An Annotation can associate a Media Element with one or more Location.. 11, 12, 13

Media Source  Source from which the respective Media Element was collected (such as CNN, DW, Twitter, Facebook...). The Media Source has among its attributes the Source Type, the Channel and the actual publisher (Actor); For social media content the actual contributors would be modelled as publishers, whereas the Channel captures the respective social media channel.. 10, 11, 17, 24, 25

Media Collection  Collection of (typically uniform) Media Elements. A Media Collection is itself also a Media Element and thus, e.g., Annotations might be associated to it.. 10, 11

Media Type  The Media Item type. Currently the model considers the following Media Types: text, video, audio and image.. 10, 11, 17, 24

Media Item  An object containing a unit of information of one Media Type such as an Audio, an Image, a Text or (a part of) a Video.. 9, 10, 11, 12, 14, 15, 24

Media Element  An information object coming from media, which is considered as one unit by the processing applied to it. A Media Element can be a Media Item, a Mixed Media Item or a Media Collection. The introduction of this more general object eases the overall modelling. A Media Element is the object to which Annotations are assigned in the project.. 5, 9, 10, 11, 12, 13, 14, 15, 17, 20, 22, 23, 24, 25

Mixed Media Item  An information object composed from several other Media Elements, e.g., a tweet containing image and text or a video composed from a number of fragments.. 9, 10, 11, 12, 14, 15, 24

MRSC  Migration-Related Semantic Concept, that is, a Concept of specific interest in the migration context, such as “racism”, “opinion former”, “refugee”, “domestic service”, “education” (see the analysis provided in Section 5 of D2.1).. 12

Observed Context/Situation  Part of the media activity and the findings generated by MIRROR analysis, which is relevant for the End User and is, therefore, observed. Observed Contexts/Situations are defined by the End User via filters and are filled by the system with relevant information; these are (annotated) Media Elements, Summaries and Findings related to the context.. 13, 14, 15, 17, 22

Organization  The representation of a real world organization.. 12, 15

Origin  The country the Media Element is originating from.. 10, 11

Perception  Perception of Europe, individual European countries and European political and societal issues in the context of migration.. 5, 9, 14, 15, 16, 21, 23

Person  The representation of a real world person or a system user. Following the ethical principles of the project, working with real names will be avoided wherever possible (i.e., via application of the Pseudonomization process, treated in Section 2.4).. 10, 12, 15, 23

Political & Societal Context  Context in the environment of the respective Actor, which influences Perception, migration decision and migration-related actions.. 15
Publication Time  The time point at which the Media Element was published/created. Alternatively, the retrieval time if the creation/publication time is not known.. 10, 11

Relationship  Representation of a relation between objects of the domain. This can for example be Media Elements or Actors.. 10, 11, 13, 25

Sentiment  A Sentiment Annotation can be used to express the prevailing sentiment polarity and stance of a Media Element content.. 12, 13

SM SubNetwork  A subnetwork extracted from a social media network. The SM SubNetwork is composed from Actors in the network, Media Elements and Relationships, representing links between them.. 10, 11, 14, 23, 25

Social Network Pattern  Perception-relevant pattern learned from the analysis of SM SubNetworks coming from one or more SM Channels.. 14

Source Type  The class or type of a Media Source. In the model we currently foresee the classes News and Social Media.. 10, 11, 24

Summary  A Summary is a piece of information derived from a Media Element (especially from a Media Collection) by summarizing it, typically using automated methods.. 5, 9, 13, 14, 22

Threat  Migration-related threat, caused by the perception and misperception of migration contexts.. 5, 9, 14, 15, 16, 17, 21, 23

Timeframe  A piece of information describing temporal information. An Annotation can associate a Media Element with a Timeframe. This time frame refers to the content of the Media Element (the time the content directly or indirectly refers to).. 12, 13

Topic/Concept  Part of the conceptualization of the considered domain. An Annotation can associate a Media Element with one or more Concepts. An important type of Concept considered in the MIRROR project is the MRSC.. 12
6 References


7 Acronyms

BPMN  Business Process Model and Notation. 17

COO   Country Of Origin. 6

MIRROR-CIM  MIRROR Conceptual Information Model. 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22

MRSC  Migration-Related Semantic Concept. 12, 24, 25

OMG   Object Management Group. 17

SM    Social Media. 14, 23, 25

UI     User Interface. 5

UML   Unified Modeling Language. 9