Co-inform

Context Matters, Your Sources Too

Co-Creation of Misinformation Management Policies

D2.1

#ThinkCheckShare

Document Summary Information

Project Title:	Co-Inform: Co-Creating Misinformation-Resilient Societies		
Project Acronym:	Co-Inform	**Proposal Number:**	770302
Type of Action:	RIA (Research and Innovation action)		
Start Date:	01/04/2018	**Duration:**	36 months
Project URL:	http://coinform.eu/		
Deliverable:	D2.1 Co-Creation of Misinformation Management Policies		
Version:	Final version		
Work Package:	WP2		
Submission date:	1/04/2019		
Nature:	Report	**Dissemination Level:**	Public
Lead Beneficiary:	University of Koblenz-Landau (UKOB)		
Author(s):	Ipek Baris, Work package leader/ Researcher (UKOB) Akram Sadat Hosseini, Researcher (UKOB) Dr. Sarah Denigris, Postdoctoral researcher (UKOB) Dr. Oul Han, Postdoctoral researcher (UKOB) Prof. Dr. Steffen Staab, Coordinator/Head (UKOB)		
Contributions from:	Martino Mensio, Researcher (OU) Orna Young, Researcher (FCNI) Syed Iftikhar H. Shah, Researcher (IHU) Dr. Somya Joshi, Lecturer (SU) Dr. Nadejda Komendantova (IIASA)		

Revision History

Version	Date	Change editor	Description
1.1	10/03/2018	All authors (UKOB)	Collaborative drafting
1.2	13/03/2019	All authors (UKOB)	Final editing and conversion to official template
1.3	25/03/2019	Lukas Konstantinou (CUT) Syed Iftikhar H. Shah (IHU)	Reviewed first draft
1.4	25/03/2019	Syed Iftikhar H. Shah (IHU)	Added 1st results of co-creation pilot session in Greece
1.5	25/03/2019	Dr. Somya Joshi (SU)	Added first results of co-creation pilot session in Sweden
1.6	27/03/2019	Lukas Konstantinou (CUT) Christiana Varda (CUT)	Reviewed 1.5 version
1.7	28/03/2019	All authors (UKOB)	Final editing
1.8	1/04/2019	Nadejda Komendantova (IIASA)	Added first results of co-creation pilot session in Austria
1.9	1/04/2019	SU	Final editing. Submission

Disclaimer

Copyright Message

Executive Summary

This report provides a first version of the ontology of the **Co-Inform** platform, the rule automation framework and the initial Co-Inform platform policies.

In Section 1, we define platform policies and legal policies, and we share the interview that we conducted with Orna Young from FactCheckNI to understand the current status of manual fact-checking. Additionally, we present survey templates for gathering initial platform policies from the stakeholders during the pilot sessions.

In Section 2 we describe scenarios which are developed from concrete examples of misinformation found on social media, and abstracted use cases for a principled development process of the Co-Inform ontology and platform.

In Section 3, we present the Co-Inform ontology, which builds on existing ontologies describing people, organizations, social media users and fact-checking. And we conclude with a taxonomy for misinformation.

In Section 4, we give more details about the structure of platform policies and platform rules. Concerning the former, we describe how, through a semantic wiki system, the policies can be co-created by Co-Inform users. Furthermore, with respect to the platform rules, we present a rule automation framework that enables the user of the Co-Inform platform to create/modify the platform rules and that executes the rules automatically when an event occurs.

In Section 5, we present initial platform policies.

We conclude the deliverable with an Appendix where we outline software technologies to be used in the rule automation framework, and feedbacks on event-condition-action rules from pilot sessions held in Sweden and Greece.

This deliverable includes inputs from WP1, WP3, and WP4. More specifically, this deliverable presents a co-creation process for gathering stakeholder requirements through WP1 to define initial policies for platform. WP2 collaborate with WP3 and WP4 to ensure the coherence of Co-Inform ontology. Policy encodings developed within WP2 will be evaluated by WP5.

Table of Contents

Table of Figures

List of Tables

List of Abbreviations

CRUD   Create, Read, Update, Delete

ECA   Event-Condition-Action

FCNI   Fact Check Northern Ireland

FOAF   Friend of a Friend

ICT   Information Communication Technology

NISRA Northern Ireland Statistics and Research Agency

SAPO   Säkerhetspolisen (Swedish Security Services)

SIOC   Semantically Interlinked Online Communities

SOP   Standard Operating Procedure

WP   Work Package

Glossary

Glossary of the deliverable	
Term	**Definition**
Scenario	A scene that illustrates some interaction with a proposed system. Also, a tool used during requirements analysis to describe a specific use of a proposed system. Scenarios capture the system, as viewed from the outside, e.g., by a user, using specific examples.
Use case	a specific situation in which a product or service could potentially be used. "there are lots of use cases for robotic hardware, from helping disabled users to automating factories"
CRUD	Create, Read, Update, Delete
Policy	Description of management and intervention in regard to openly created user content and user behaviour on an interactive online platform
Rule	Machine readable and (partially) automated policies.
Ontology	An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them.

1   Introduction

Within Co-Inform, the WP2 aims at empowering different stakeholders by enabling resilience to misinformation and cultivating informed behaviours and policies. For this concerted aim, WP2 will provide policy encodings and work in tandem with the ICT tools and services that will be developed by WP3. By itself, WP2 will build new ontologies and develop available ones for modelling misinformation, disinformation, fake claims, rumours, veracity, social and information networks. Further and by interlinking with WP3, the WP2 will facilitate the mining of misinformation and its validation with corrective information from fact checkers in the form of a browser plugin, repository of news and rebuttal from the community (e.g rbutr tool[1]), and a dashboard.

In this deliverable, we propose an ontology-based framework as a support system for human decision makers, particularly for three groups of stakeholders: citizens, journalists and policy makers.

1.1   Objective of WP2

Within the objective of **Co-Inform**, which is to create a collaborative platform on which we manage misinformation, WP2 is tasked with the objective of defining a) the collaborative aspect of this platform and b) the process of managing misinformation. This objective renders the question: What does it mean to *manage misinformation*?

Regarding this specific objective, the idea of "managing misinformation" is currently not supported by **standard rules, policies, or recommendations** for how to tackle (thus how to "manage") online misinformation.

Addressing this, WP2 aims to identify standard policies for managing misinformation and to encode a part of them as a) **machine-readable, automated sets of platform rules and b) non-automatable platform policies that are performed by humans.** The resulting policies will be used to guide the development of misinformation services of WP3 and their deployment in WP4. Further, WP2 also focuses on defining a set of interventions which will be integrated in WP4, while their impact will be tested in WP5.

WP2's objectives therefore encompass the merging between a co-creation approach and key indicators for handling misinformation, which requires that we provide initial sets of

---

[1] **http://rbutr.com/, retrieved 25.03.2019**

management and intervention policies for guiding the platform functions, and also that we encode them in machine-readable format.

To summarize, main objectives of WP2 are to:

- Co-create key indicators for handling known misinformation
- Identify management policies to guide Co-Inform misinformation handling processes and implementations
- Identify a set of intervention policies
- Provide encodings of management policies for analysing and developing Co-Inform

1.1.1 Objective of Deliverable Task 2.1

In this deliverable "Co-Creation of Misinformation Management Policies", we make a key distinction between policies and rules that will resonate throughout the document:

Definition: Policies are described by human readable text, while rules are machine readable and (partially) automated policies.

The term "policy" encompasses:

- Platform policies: descriptions of what users can/cannot do on the **Co-Inform** platform (Section 0)
- Legal policies fixed by lawmakers (Section 0)

The description of work of WP2 is entirely focused around platform policies and platform rules.

For accomplishing WP2, our methodology contains conceptual parallels to software development, with the key difference of allowing for the co-creation of platform policies and platform rules. This feature, the platform policies and rules co-creation by the community participants, is what renders the platform more effective and more efficient.

An important constituent of such flexibility is the definition of an ontology, in Section 3, upon which policies and rules will be defined. In a nutshell, an ontology is a way to represent our knowledge on a specific topic, defining concepts and the relationships between them. The ontology is therefore the essential substrate for the definition of platform policies as it is, effectively, a representation of the platform ecosystem.

The ontology defined here will be collated with other ontology parts coming in other work packages. In this context, the Deliverable Task 2.1 of WP2 defines core entities needed in WP2, and therefore presents an ontology and core framework for WP2.

1.1.2 What a policy is

A policy is "explicit and implicit norms, regulations, and expectations that regulate the behaviour of individuals and interactions between them" [Butler, 2007]. Policies can be symbolic or can reflect an action, or both. In Co-Inform project usually second case is considered; but the first type is also needed to coordinate the behaviour of the characters involved in the Co-Inform system. Policies can be seen from several facets:

- Rational efforts to organize and coordinate
- Constructions of meaning & identity
- External signals
- Internal signals
- Negotiated settlements and trophies
- Control mechanisms

This deliverable rests on the premise that two levels of "policy" exist, which are mostly distinguished by technical differences. The first sense of "policy" is **platform policies** which are constrained within the functions of an online co-authored and co-managed platform (such as the **Co-Inform** platform as virtual infrastructure and ontological framework). Such policies comprise sub-systems of automated rules that are created for regulating standard functions and special functions triggered by conditions that have been inputted prior. **This technological and platform-oriented definition of "policy" is the main subject of this deliverable and of WP2.**

The second sense of "policy" is **legal**: this sense is socially more common and applies to the institutional infrastructure of legal frameworks, such as EU policies in the supranational and intergovernmental design of the European Union.

Below sections outline both senses of policy ("platform" as in distinct from "legal") and conclude by drawing parallels as future outlook for development and application beyond WP2 and the self-contained aims of **Co-Inform** as a European Union project.

Platform Policies

"Platform Policy" relates to the platform-internal regulations of the final **Co-Inform** product, which engages users by providing them with relevant functionalities. Therefore, a productive starting point for designing platform policies is to utilize existing platforms and online services as reference, by identifying essential parallels in basic principles.

The following table shows possible cases in the typical operation within a fact-checking scenario of the **Co-Inform** platform. The cases are phrased as general domains that subsume a multitude of different types, conditions, and events that fall under the particular policy. Below presented cases are of initial form and derived tentatively from Wikipedia, which we utilized as (initial) reference platform (Table 1).

Table 1. Possible list of existing platform management policies			
Policy	**Possible characteristics of the case**	**Action**	**Medium**
Hate speech	Pages that disparage, threaten, intimidate, or harass their subject or some other entity, and serve no other purpose	Deletion	Wikipedia
Incorrectness	If someone believes a poster interpreted the consensus incorrectly	Review needed	Wikipedia
False rumour	If there were substantial procedural errors in the deletion discussion or speedy deletion	Revert	Wikipedia
Unreliable content	If articles that cannot possibly be attributed to reliable sources including neologisms, original theories and conclusions, and articles that are themselves hoaxes (but not articles describing notable hoaxes)	Deletion	Wikipedia
Restricted content	If the article contains possible restricted content (posted) by user	Account suspension	Reddit
Thread trolling	If someone purposely posts something controversial or off-topic messages in order to upsetting people and get a rise of other people	Warning	Reddit
Restoring Suspended Account	After an SOP implementation within stipulated period of time the suspended account will be re-activated	Account Re-Active	Google

Legal Policies

In the context of WP2, "legal policies" simply refer to policies informed by existing laws and regulations. For instance, a policy forbidding to post criminal content mirrors and enforces a law against such content. This example also clarifies that, between platform and legal policies, it exists an overlap as the former are to be shaped, of course, in compliance with the existing legal framework. However, platform policies are a broader ensemble: as a platform policy, it might be envisioned, for example, to restrain the use of certain words, like curse words, for the sake of maintaining a discussion civil, in spite of such vocabulary not being illegal per se. Moreover, this type of policy is a fruitful prospect for future development and research. Upon completing the **Co-Inform** platform and its core ontology, we foresee **a novel avenue for aiding policymaking that is realistic for the parameters of the Internet age:** By amplifying implications from machine learning the behavioural patterns around misinformation handling online, **big data can be intelligently leveraged to "learn" effective measures, and thus to transpose platform policy to legal policy.**

1.2 Current status of manual fact-checking

We conducted a brief interview with a small-sized and Northern Ireland-based fact checking organization, FactCheckNI, who is also a **Co-Inform** project partner (time of conduct: January 7th, 2019. Interview partner: Orna Young, FactCheckNI (FCNI hereafter)).

Our questioning revolved around daily fact-checking tasks of incoming claims management and corrective intervention in the case where misinformation can be identified. This brief survey aimed to establish a realistic impression of the **scope and challenges of manual fact-checking** operations where no automation is involved. It further intended to gain **a first look at potential areas where automation can be effectively integrated** to produce the double outcome of **a) accelerating fact-based operations and b) amplifying the reception of benefits by audiences.** Based on the following material, a first-hand impression is gained regarding the general standard framework that underlies manual fact-checking operations.

In the case of FCNI, **the general principle is that manual fact-checks are prompted by external request from users, instead of decided at will.** The vast majority of FCNI's claims are sourced from news media/social media, not directly from contributors. As an organization, FCNI does not fact-check general areas or topics by own decision. FCNI only fact-checks specific claims made by public figures/ representatives/ organizations. Those submitted by members of the public are subject to assessments as covered below (Table 2).

Table 2. Assessment table for manual claim selection (Case of FactCheckNI)	
Questions	**Answers**
Before fact-checking	
Is the claim currently in public discourse?	Issues/claims which are receiving coverage in the media; and or has generated a lot of discussion.
Is it about an issue relating to Northern Ireland specifically?	FCNI does not factcheck issues/claims outside of Northern Irish jurisdiction.
Has claim been made in a public forum? (i.e. on the record, in a public statement, or published by an individual or organization)	FCNI does not factcheck claims made by individuals in private conversations/ on personal and private social media pages unless without their explicit permission.
Who posts the claim and how?	for example, if it is a member of a political party, we will have to ensure we have not focused specifically on fact-checking on political party in particular. If it is a claim that is submitted to us, we must ensure that the source is accurately representing the information they are giving us.
Which type of users post claims?	The general public, those with specific socio-political interests (e.g. the environment).

During fact-checking	
Where does evidence come from?	We are affiliated with the Northern Ireland Statistics and Research Agency (NISRA) – Similarly, once the data has come from academic/reliable sources – such as ONS, or other traceable organizations/sources we will employ (or indeed, challenge) them. This frequently means we draw on a range of sources for the same topics to draw comparisons or reveal any inconsistencies.
After fact-checking	
When you have fact-checked claims, based on which criteria do you post checked claims on your website?	If we have decided to proceed with a factcheck we will have already decided that it is suitable for publication. Those issues/claims that are made and cover thematic areas / issues which cannot be fact-checked (such as future implications of a particular issue) will sometimes be written up in to a blog post--discussing the facts relating to the thematic area, but not making any judgement on the facts relating to the future implications of the topic.
Which topics do you most frequently receive as claim?	We have had calls to factcheck issues relating to Brexit (much of which is future dependent, thus impossible) or relating to the activities of political parties in Northern Ireland. Those fact-checks that we have published previously on these areas have tended to be out most shared.

1.2.1   Surveying focus groups

As part of the co-created development principles of Co-Inform, the pilot group conducts intensive workshop sessions with focus groups (see the document "Co-Creation Workshop 1") that are composed of sample representatives for diverse stakeholders (according to the "Data Collection Framework" of the WP1/WP5 document titled Co-Creation Workshop 1). This evaluation framework provides a common methodology that can be applied to all co-creation workshops across the different locations, but allows for customization of surveying elements, depending on local implementations or the goals of survey. This flexibility is integrated in the design of the deliverables of WP1. A focus group is typically composed of the "stakeholders" (e.g. future users of the platform) which are defined here as citizens, policymakers, and journalists. In other cases, a focus group can be composed of project partners. This survey material can therefore be applied to all types of co-creation workshops and be integrated in their data collection.

For the WP2-specific aim of designing a core ontology that lies behind the future Co-Inform platform, we contribute dedicated survey material that is designed to derive a "wish list" of platform functions and mechanisms from the stakeholders. More specifically, the material encompasses:

1. **Two sets of questions**:
   a. one set that pertains to the desired features automation that WP2 will integrate in the core ontology,
   b. and another set that relates to essential mechanisms of trust.
      - It is particularly this latter set of questions that address how to embed incentive and motivation to disseminate checked facts, which is a feature that is able to set apart Co-Inform as a platform from existing fact-checking websites and browser plug-ins.
      - For this aspect therefore, we will draw reference and precedence from online expert communities and the role of experts in fact creation, maintenance, and visibility.
      - For example, the involvement of topic experts plays a crucial role in a) generating trust in a checked fact for the news consumer and b) the consumer's willingness and motivation to share and disseminate the fact that has been checked.
2. A template that is to be filled out by focus group participants. This template is held in the rules grammar of "Event-Condition-Action" and embodies output that can be utilized by WP2 instantly by yielding scenarios and, subsequently, use cases.

As described above, we present below the two-part material for surveying essential information that is required by **WP2** to ensure **close internal collaboration and external success of Co-Inform** (Table 3, Table 4). Additionally, we provided example answers and a questionnaire to survey the news categories of most interest to the participants (Table 5, Table 6).

Table 3. Question sheet for stakeholder surveys
Which focus group do you belong to: ☐ Citizen ☐ Policymaker ☐ Journalist
a. **Questions that will be used to build the automatic system behind the Co-Inform platform**
1. <u>Which news topic</u> would you like to not miss if you had an automatic news alert system?
b. **Questions that will be used to leverage mechanisms of trust in the Co-Inform platform**
2. Under which conditions would you automatically **trust** online news content that is <u>shared by a friend</u>?

3. What would make you **trust** a <u>media source</u>?
4. What would make you **trust** a <u>fact-checker</u>?
5. What would make you **trust** an <u>opinion</u>?
6. What would make you **trust** a <u>journalistic piece</u>?
7. What would make you **trust** a <u>fact that has been checked by a fact-checker</u>?

Table 4. Template to be given to stakeholders
Are you a: ☐ **Policymaker**   ☐ **Journalist**   ☐ **Citizen**
<u>Read this scenario first:</u> Imagine an online platform that you can fully and freely customize, regarding 1.   Which kind of news you see first 2.   Which kind of news to flag as potentially false 3.   Which kind of news to submit for fact-checking 4.   To whom to submit news for fact-checking 5.   Everything else in between Your answers will be used for automating high-tech functions that would be important to you, <u>but which do not exist yet in any online service that you know.</u>

Event	*Please state any situation that relates to misinformation that you face when using the Web (example: Facing fake facts in a news article about societal groups of specific nationality)*
Condition	*Please state any condition of the above situation that causes you to act or demand action* *(example: Nobody in the replies has corrected the misrepresentation of specific nationality)*
Action	*Please state any action that you would enact yourself or demand from administrators* *(example: Request deletion. Alternative example: Flag news article as potentially false.)*

Table 5. Example answers to the template to be given to stakeholders		
User	**Policymaker**	**Journalist**	**Citizen**
Event	Facing a potentially misinforming article in web	Facing a news on web about a terrorist attack in his/her country	Facing a post that has attracted attention and actively shared in social media
Condition	There is no corrective information available from fact checkers about this article	News contains personal information like address and phone	Post contains harassment of specific party group

		number of family of possible terrorist	
Action	Submitting the article to fact-checkers for validation	Editing partly misinformation news to the correct news	Receiving a notification

Table 6. Examples of news topic	
Instructions: For question 1 of Table 1a, please answer with your desired categories, together with example topics that belong to that category. Example: I am interested in getting news about the EU, and specifically on the topic of Brexit.	
Example category:	**Example topics:**
EU	Brexit, Immigrants
World Politics	US, EU, Asia,
Health	Research in diabetes, cancer etc.
Environment	Climate change, nuclear plant construction, pollution, animal extinct
Economy	Finance, trade, bitcoins, business

2   Scenarios and Use Cases

This section follows the tried and proven convention of software engineering design that **envisages realistic scenarios** and derives **general use cases** to reflect the system of aims within the product to be developed. In the software design context, **a scenario is**:

- A scene that illustrates some interaction with a proposed system
- A tool used during requirements analysis to describe a specific use of a proposed system.

Therefore, scenarios capture the system, as viewed from the outside, e.g., by a user, using specific examples. In this section, two scenarios will be outlined as a first basis for conceptualizing and describing the core structure of the platform.

Following the scenario, an abstracted **use case** will be formulated to describe an in-built functionality of the **rule automation framework integrated to the Co-Inform platform**.

In what follows **we present three scenarios, with the relative use cases,** that concern the rules triggered by misinformation **on social media.**

2.1   Scenario A: User creates single rule as part of wider policy

Henry is a registered user in the Co-Inform platform who has an interest in preventing misinformation. He has only recently joined the platform and has not set any preferences in his user profile regarding preferred topics or types of misinformation. However, he would like to be notified whenever a general piece of misinformation, one that is currently being actively shared in social media and attracting attention, is detected by the platform.

Therefore, he customizes the detection function of the platform by writing a rule about notification. The single rule that Henry creates is part of a wider policy (receiving notifications about instances of misinformation) within the platform; his created rule follows the structure below and contains conditions that trigger an action that Henry desires.

Policy 1: Receiving notifications about instances of misinformation

- **Rule**
    - **Event:** Detection of misinforming post
    - **Condition:**
        - Post contains harassment of specific party, group
        - The poster is likely to share misinformation (based on language patterns and other components that the platform detects as likely misinformation).

      o   **Action:** Subscriber of the platform is immediately notified

After Henry has entered this rule, he receives a first notification. The suspicious post in question is the following Facebook post (Figure 1)

Figure 1. Facebook post containing a false claim

In other words, Henry has been alerted about a post that contains the following claim, which has been detecting by the platform as being highly possibly false (Figure 2):
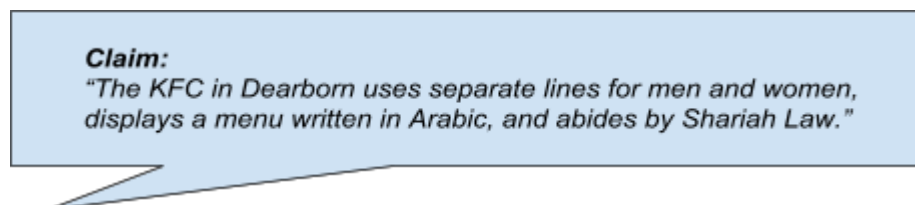

Claim:
"The KFC in Dearborn uses separate lines for men and women, displays a menu written in Arabic, and abides by Shariah Law."

Figure 2. Claim indicated in the post

Thus, Henry's rule has notified him about a detected misinforming post that is shared and is gaining popularity on Facebook at the moment.

2.2 Use Cases for Scenario A

The above scenario can be modelled as two use cases. First use case specifies how Co-Inform users create platform policy for setting up notifications and other one specifies that when the system receives input of the event, how the rules of corresponding policy are executed by rule automation framework (Figure 3).
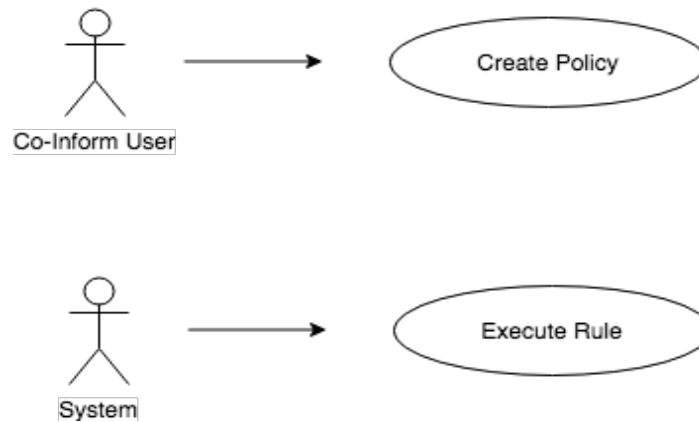

Figure 3. Use case diagram for the scenario A.
Henry is Co-Inform user, and system is the rule automation framework.

2.2.1 Use case A.1: User creates platform policy

Actor: Co-Inform User

Flow of events:

1. Actor adds the policy and corresponding set of rules by using rule editor.
2. Rule manager receives the rules as input from rule editor and converts them to rule language that is accepted by rule reasoner.
3. Rule engine checks the rules whether there is conflict with core ontology, and then if there is no conflict, rule manager registers the rule.

Preconditions:

User must be registered.

2.2.2 Use case A.2: System executes platform policy

Actor: System

Flow of events:

1. Rule automation framework receives the signal from misinformation detection module of WP3.
2. Rule manager retrieves rules corresponding to the event in the repository and loads them into rule reasoner.
3. Rule engine executes the rule and returns the action.

Recalling the policy defined by Henry, above flow is repeated two times for, as an example, the hate speech model and by checking the history of the poster. Firstly, the rule automation framework triggers hate speech module. Secondly, after receiving input from the hate speech module, the system triggers the module which checks the history of the poster. Finally, if all conditions are satisfied, rule reasoner executes the final action which is to notify Henry.

2.3 Scenario B: User submits an article for fact checking

Ali is a policymaker and a user of Co-Inform. One day he receives an alert that a potentially misinforming article is published, he finds that no corrective information is available from fact checkers about this article. He decides to submit it to Full Fact for validation through endpoint of the system. Co-Inform applies following policy in order to approve the submission and send it to fact checkers.

Policy 2: This policy describes how users/policymakers can successfully submit an article to Full Fact for validation and Co-Inform send it to fact checkers.

For this purpose, Co-Inform should check the following conditions:

- No other user has already submitted article to Full Fact for validation. If it has been submitted in past by another user, submission request should be deleted, and user is warned and also tagged to receive notification when fact checking organization publish article validation result.
- Also, an article will be sent to fact checkers when specific time has passed from its publishing time. If this condition is not met, the request must be sent to waiting list and system warn the user.
- **Rule 1**
  - **Event:** A user submits an article to full fact for validation.
  - **Condition:**
    - No other user has already submitted article to full fact for validation.

- When user register his request (submits article), x time has been passed from published time of article.
    - o **Action:** Send article to fact checkers/add article to waiting list, notify user, tag the article, etc.
- **Rule 2**
    - o **Event:** A report is published about a news article.
    - o **Condition:**
        - Article is submitted for fact checking through policymakers.
        - Article not deleted yet from publisher media.
    - o **Action:** Send notification to submitted user, add user and article in the list of submitted articles for fact checking.
- **Rule 3**
    - o **Event:** When time to send an article in waiting list to fact checkers comes. (Articles in waiting list are those that did not meet time conditions and have not been sent to fact checker in submitted time).
    - o **Condition:**
        - There is no report and validation result about article from fact checkers side.
    - o **Action:** Send article to fact checkers, notify user, tag the article.

2.4 Use Cases for Scenario B

The *scenario B* that we explained in previous section can be modelled as two use cases. First use case specifies how Co-Inform user (Policymaker) send submitting request to fact checking and second one specifies that when the Co-Inform system receives request for fact Checking, how the rules of corresponding policy are executed by the framework (Figure 4).
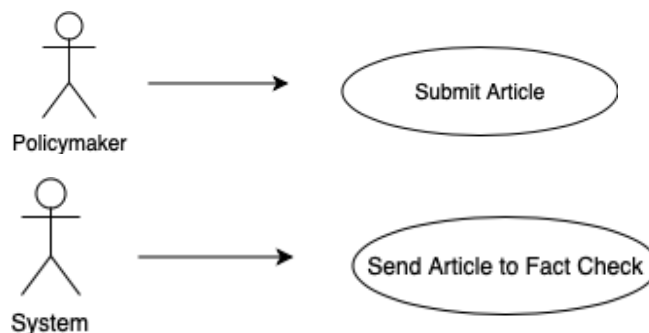

Figure 4. Use case diagram for the scenario B.
Ali is Co-Inform user who is a policymaker, and system is the rule automation framework.

2.4.1   Use case B.1: Policy maker submits an article

Actor: Co-Inform User (Policymaker)

Flow of events:

1. Policymaker submit an article for validation.
2. Rule mapper convert *submitting* command to corresponding rules language that is accepted by rule reasoner.

Preconditions:

- User must be registered.
- The role of user in platform must be policymaker.

2.4.2   Use case B.2: System executes platform policy

Actor: System

Flow of events:

1. Rule automation framework receives the signal from misinformation detection module of WP3.
2. Rule manager retrieves rules corresponding to the event in the repository and loads them into rule reasoner.
3. Rule engine executes the rule and returns the action.

2.5 Scenario C: Semi-supervised content flagging

There might be cases where fully automated rules cannot be applied. In these cases, user engagement should be required in order to integrate a **semi-automated element in form of human judgement**. For example, the following can be a policy:

> *"Misogynistic comments should not be allowed to be published. At first time, an offender will receive a warning, but at second time the offender will receive a severe warning alongside the temporary freezing of his account. Finally, at the third time, the offender's account will be deleted".*

In these cases, a comment can be flagged by anyone as a candidate for a misogynistic event. A flagged comment is classified as a misogynistic offense eventually if an editor agrees on it being misogynistic. The following Policy 3 and Policy 4 are illustrated (Figure 5).

Policy 3: This policy describes how a comment is classified as a misogynistic if at least one editor confirms it.

- **Rule**
  - **Event:** A user tags a comment as misogynistic
  - **Condition:**
    - No condition.
  - **Action:** Send tagged post to editors in order to verify it. (The subsequent decision by the editor represents a semi-supervised element.)

Policy 4: This policy describes how to deal with a user who posts misogynistic comments. First time the offender will receive a warning, second time the offender will receive a severe warning together with a temporary freezing of his account. Finally, after third time the offender's account will be suspended.

- **Rule 1**
  - **Event:** A comment is recognized by the editors as a misogynistic.
  - **Condition:**
    - Account is not in offender list.
  - **Action:** Send warning to account, add account to offender list.

- **Rule 2**
  - **Event:** A misogynistic content is posted by an account that is in offender list.
  - **Condition:**
    - Account not received a severe warning previously.
  - **Action:** Send a severe warning to account, temporary freeze the account, tag account as severely warned

- **Rule 3**
  - **Event:** An account that received a previous severe warning posted misogynistic content again.
  - **Condition:**
    - No condition.
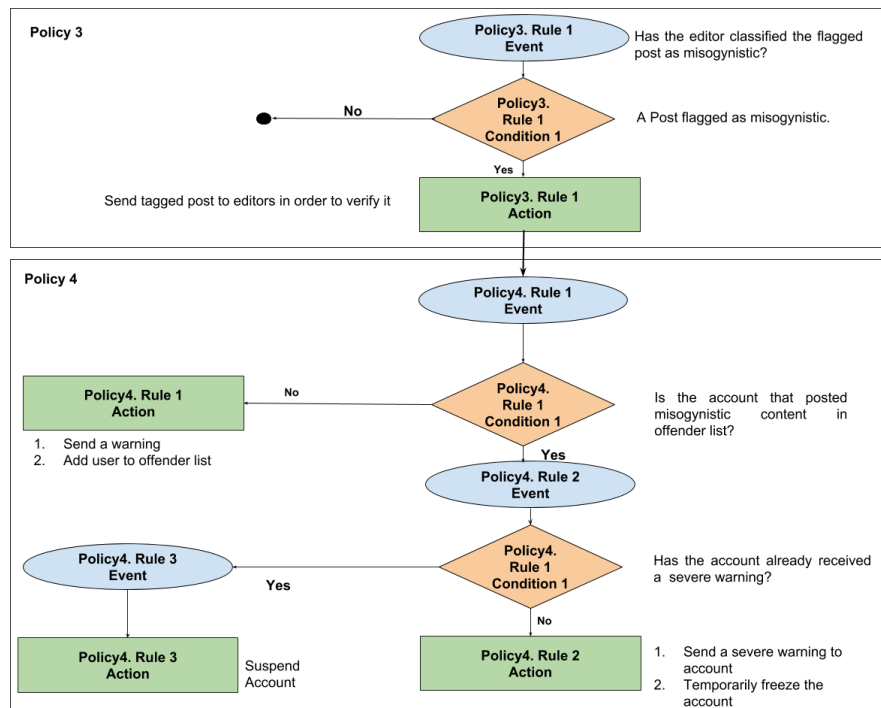 - **Action:** Delete related account.

Figure 5. A system of semi-supervised rules in form of a scenario

In this kind of semi-supervised policy scenario, the human intervention by manual judgement becomes necessary due to the ambiguous nature of many controversies that arise from political nature of social norms and moral standards. As result of judgement and possible discussion, a policy may even have to be revised. The Co-Inform platform allows for this possibility in order to ensure a sustainable policy system and rule framework, as described in Section 4.1.

3 Co-Inform Ontology

This section outlines how the platform elements in the Co-Inform ontology are grounded and combined. An ontology in computer science is a formal representation of knowledge with the concepts, their properties, and the relationships between the concepts in the domain. Ontologies provide common understanding of the structured information among people and software agents [Noy, 2001], and management on the relevant data. They are widely used to model the domain of applications such as search engines, e-commerce, multi-agent AI systems, etc. By posing as underlying framework, ontologies are a promising instrument to overcome the common challenges between different management domains. For this reason, the WP2 ontology stands in close relation with, and exists in technical collaboration with WP3, WP4, and WP5 of Co-Inform.

In general, the development of ontology includes the following tasks [Noy, 2001]:

- defining classes in the ontology,
- arranging the classes in a taxonomic (subclass–superclass) hierarchy,
- defining properties and then describing restrictions

As shown above, the final ontology will follow the same overarching grammar, but everything else will differ according to the subject domain of the ontology. For instance, in the case of WP2, the ontology will propose to describe rules for management of misinformation, as well as intervention methods, in Co-Inform. Since we exploit our analysis of misinformation in the particular field of social media, we need to use models that represent this setting. The ontology need not be developed from scratch, but needs instead to utilize related vocabularies to define core entities such as post, actor, etc.

Fortunately, we can use the following existing models for our purpose: among them the FOAF, SIOC, and Schema.org ontologies [Fernandez, 2014]. Firstly, FOAF (the Friend Of A Friend vocabulary) describes people, their properties, and the social connections of people who are linked across social web platforms. Secondly, SIOC (Semantically Interlinked Online Communities) originates from modelling discussion boards and is used today to model users and social interactions, as well as content and the reply chain where this content is embedded. This ontology re-uses classes and relations of FOAF. Lastly, Schema.org provides vocabularies that are agreed upon by the major search engines (Google, Bing, Yahoo!), and capture knowledge about people and social networks. "Vocabularies" here contain aim-oriented catalogues of tags for defining item types (Person, Place, Organisation, Review, Event, etc.) and social relations (knows, colleague, children, parent, sibling, relatedTo. etc.).

Below Figure 6 presents how the above ontologies integrate in conjunction, which enables us to capture actions and interactions of the user within an online community [Fernandez, 2014]. The online community provides information about the social network of the user and the content she produces.
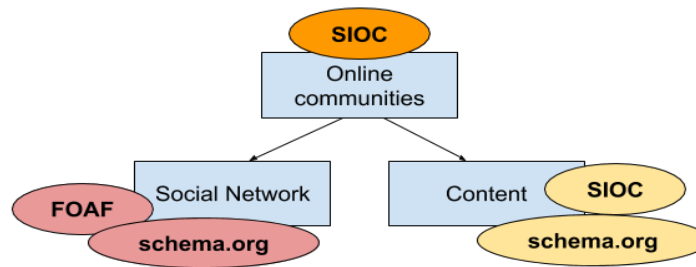
Figure 6. Existing ontologies in the context of online communities

3.1 FOAF

The name 'FOAF' is derived from traditional internet usage, an acronym for 'Friend of a Friend'. The reason for choosing this name was to reflect the concern with social networks and the Web, urban myths, trust and connections. FOAF collects a variety of terms; some describe people, some groups, some documents. Different kinds of application can use or ignore different parts of FOAF.

FOAF is a project devoted to linking people and information using the Web. Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked. FOAF integrates three kinds of network: *social networks* of human collaboration, friendship and association; *representational networks* that describe a simplified view of a cartoon universe in factual terms, and *information networks* that use Web-based linking to share independently published descriptions of this interconnected world. FOAF does not compete with socially-oriented Web sites; rather it provides an approach in which different sites can tell different parts of the larger story, and by which users can retain some control over their information in a non-proprietary format. FOAF fits perfectly the requirement of representing the social components of our ontology. Social connections of different users in FOAF is made by *foaf:knows* property. FOAF depends heavily on W3C's standards work, specifically on XML, XML Namespaces, RDF, and OWL. The following diagram[2] uses the example of foaf:Person to illustrates how the properties of classes justifies a diverse range of connections (Figure 7).

---

[2] **https://www.slideshare.net/Channy/modeling-user-interactions-in-online-social-networks/8-FOAF_Ontology_describing_persons_their**
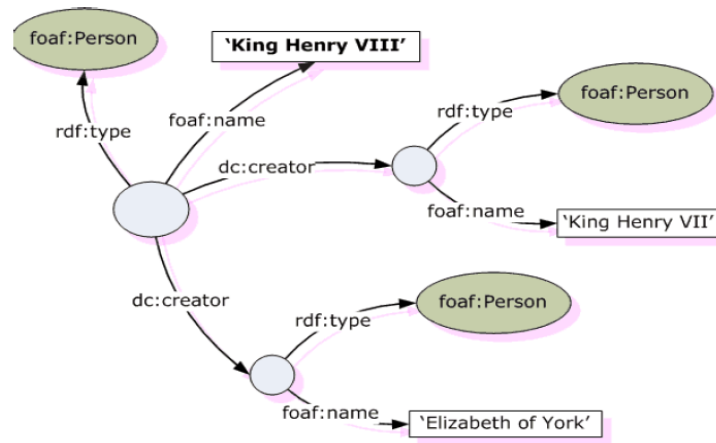
FOAF

Figure 7. Structure of FOAF by example of foaf:Person

3.2 SIOC

Online community sites are like scattered islands, you may find some information in a site, but not know that there are missing pieces of relevant information on other forums. The name "SIOC" is an acronym for "Semantically-Interlinked Online Communities". SIOC Core Ontology was developed to model the main concepts and properties required to describe information from online community sites (e.g., message boards, weblogs, wikis, etc.) about their structure and contents, and also to find new information and connections between contents and other community objects. The SIOC Core Ontology definitions have been written by combining a computer language (RDF/OWL) that makes it easy for software to process some basic facts about the terms in the SIOC Core Ontology, and consequently about the things described in SIOC documents. Since Co-Inform is aiming at co-relating different sources of information to support claims for fact-checkers in their verification process, this model seems very relevant, and we will use it for our ontological framework. We selected SIOC, because it is a generic ontology that is not designed for any specific social media platform.

Figure 8 shows the overall structure of the ontological model of SIOC. Considering that *User Account* is introduced as class and that it is clearly connected to the different types of *Post* that are being produced in different societies is essential to the purposes of making sure if a claim published in social media or online news media. By accepting SIOC we are associating different types of misinformation classified in the Co-Inform ontology to specified types of users and also to distinguished types of social platforms. This can help to show different views on statements and claims, dependent from where and from whom they originate. Additionally, by using SIOC ontology, we can also model our users in Co-Inform platform.
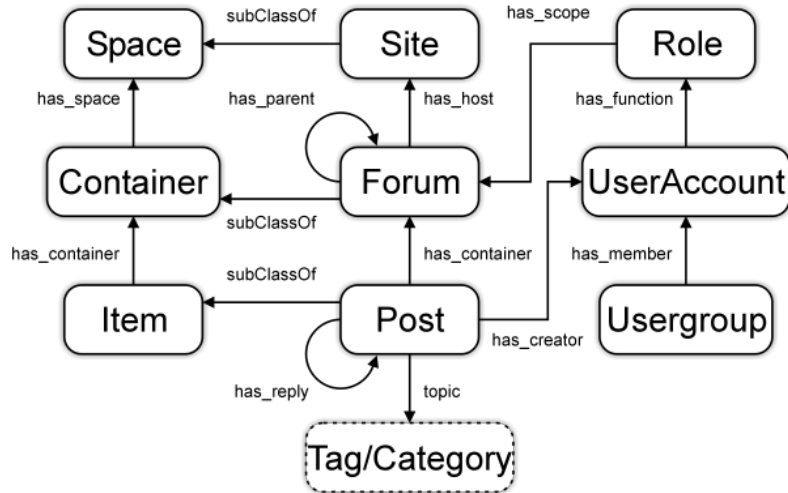
Figure 8. Graphical overview of the SIOC Core Ontology

3.3 Combining FOAF and SIOC for user relationships

From the above overview of FOAF and SIOC, we identify key classes that are the focus of these ontologies and describe concepts for particular domains. The combination of classes across the two ontologies is commonplace in order to capture different Web ecospheres realistically. A visual illustration is shown in Figure 9:
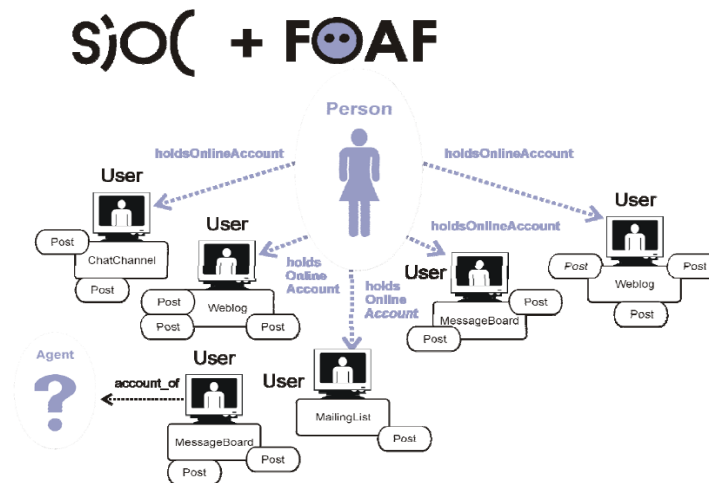

Figure 9. Combining SIOC and FOAF[3]

Therefore, we arrive at the following table of key classes that are relevant for the implementation of **Co-Inform (**Table 7**):**

Table 7. Key classes from SIOC and FOAF for the Co-Inform platform		
SIOC	Class: sioc: **Post**	*Post* describes messages, claims or articles posted by a User to a Forum, which is a subclass of sioc:Item, and foaf:Document.
	Class: sioc: **Role**	*Roles* represent functions or access control privileges that Users may have within a scope of a particular Forum, Site, etc.
	Class: sioc: **User**	The *User* is a subclass of *foaf:OnlineAccount*, which describes properties of an online account.
FOAF	Class: foaf: **Agent**	An *Agent* can be a person, group, organization or software bot. In summary entities that can perform actions. The subclass that representing people is *Person*. It also has two other subclasses, *Group* and *Organization*.
	Class: foaf:**Person**	This class represents people, a subclass of *Agents* in FOAF. A person will normally have a User account on a site and will use this account to interact within the community and create new content.

3.4 Schema.org vocabularies for fact checking

Schema.org[4] is a collaborative, community driven initiative to provide a single schema covering wide range of vocabularies for structured data on the Internet, on web pages, etc. Schema.org has been promoted by major search engines such as Google, Yahoo, etc and now it is widely used as a common vocabulary by over 10 million websites to mark-up their web pages and email messages. For instance, ClaimReview[5], CreativeWork[6] and Rating[7] are schema.org vocabularies to annotate fact checking claims and are used to show fact checking claims in search engine results[8]. Schema.org is constantly being developed for, and in collaboration with, fact-checking and trust-building endeavours.

---

[4] **https://schema.org/**
[5] **https://schema.org/ClaimReview**
[6] **https://schema.org/CreativeWork**

[7] **https://schema.org/Rating**
[8] **https://developers.google.com/search/docs/data-types/factcheck retrieved on 07.01.2019**

How to understand and read schema.org vocabularies is simple in principle: The data types as defined by schema have multiple possible properties that are factually used on the Web, which in sum result in a grammatical structure made of schema vocabularies.

Each type within any schema.org vocabulary is a data type that appears as "instances" with values for specific properties. At the time of writing, for instance, the data type Text has 457 properties at schema.org, including the following:

- articleBody (for news articles)
- caption (for images or videos)
- citation (for referred publications, web pages, scholarly articles etc.)
- commentText (for user comments)
- and many more.

Based on the broad usability of Schema.org as industry standard and established system of annotation, we select requirements for fact checking. As result, the below Table 8 shows a selection of relevant data types for Co-Inform.

Table 8. Schema.org vocabularies for the Co-Inform platform	
Type: ClaimReview	
Property	**Description**
claimReviewed	a short summary of the specific claims being evaluated.
reviewRating	assessment of the claim and is a subclass of schema Rating.
url	link to the full article of the fact check.
author	publisher of fact checking article and is a subclass of schema Organization.
datePublished	date when fact checking article is published.
itemReviewed	describe the claim being made and is a subset of CreativeWork.
Type: CreativeWork	
Property	**Description**
author	publisher of the claim and is a subclass of Person or Organization.
Type: Rating	
Property	**Description**
alternateName	truthfulness rating of the reviewed claim in textual format. For example, "Mostly True", "False".

3.5 Core Ontology for Co-Inform

Finally, having grounded and combined the most relevant elements from FOAF, SIOC, and Schema.org as shown above, we can derive the following outline of a core ontology that operates essential elements of **the Co-Inform platform**, as shown in Figure 10:
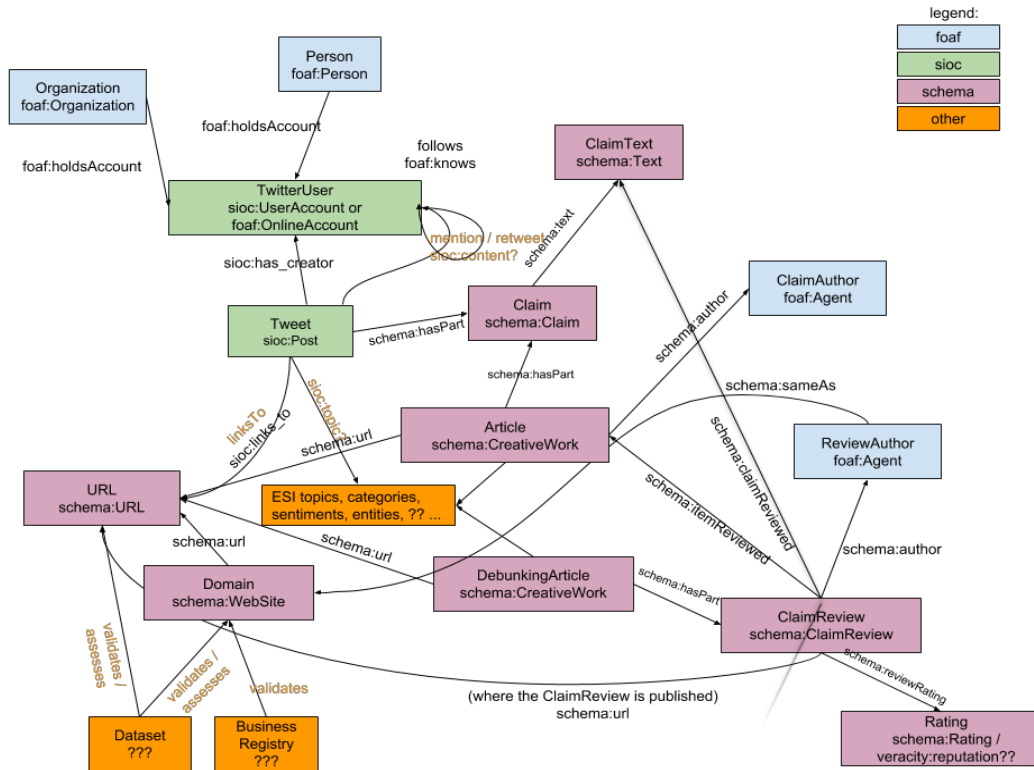

Figure 10. Core Ontology of Co-Inform

The following Table 9 shows the example classes of the Co-Inform ontology. **The relationships between classes** are not further elaborated in this deliverable, but will be of inevitable focus in later deliverables, and will be the subject of collaboration with subsequent Work Packages.

Table 9. Classes and examples of Core Ontology elements			
Ontology	**Class**	**Based on**	**Examples**[9][10][11]
FOAF	ClaimAuthor	foaf:Agent	Person, people, organization, group, or bots.
	ReviewAuthor	foaf:Organization	Fact-checkers
	Person	foaf:Organization	Social institutions, companies, societies etc.
		foaf:Person	Person that is alive, dead, real, or imaginary
SIOC	TwitterUser	sioc:UserAccount	Online account of a member of an online community; its activity is connected to items, posts, forums, sites
	Tweet	sioc:Post	Article or message posted by a UserAccount to a forum, common subject conversations, reply chains, attached files
Schema	Claim	schema:text	Publisher of the claim, truthfulness rating, short summary of claim, name of fact-checking organization
	Article	schema:CreativeWork	Articles or news articles, video, blog post
	URL	schema:URL	Link to the page hosting the claim to be fact checked, or of the fact check itself
	Domain	schema:WebSite	The homepage of the organization that is making the claim
	DebunkingArticle	schema:CreativeWork	Counter-claiming articles or news articles, video, blog post
	ClaimReview	schema:ClaimReview	fact-checking summaries of news articles or claims, TV/Radio, broadcast, media
	Rating	schema:ReviewRating	"True" or "Mostly true", numeric ratings

---

3.6 Misinformation Taxonomy

In this section, we describe the labels which will guide the classification of Misinformation. To this end, we blended together different approaches, such as the FirstDraft as well as Snopes and Opensources classifications.

We thus have two dimensions that quantify, respectively, the truth rating and the intent to harm, and a third one, in case the piece of information is not verifiable, thus not being false or true a priori. In the macro-classes defined by those dimensions, we then include more fine-grained tags, as illustrated by the table below (Table 10):

Table 10. Misinformation Labels Organization		
Labels	**Harmful**	**Not Harmful**
True	Personal Information, Leaks	
False	Misinformation: Fake News, Conspiracy Theories, Junk Science	Satire
Not Verifiable	State News, Extreme bias, opinions distorted as facts	Unproven scientific theories, Opinions

In the table, the second level of fine-grained labels is more context dependent, requiring both automatic detection and human judgement. To complete this characterization, it is worth noting that the above labels could be complemented by the label hate speech: indeed, such register, in spite often being a signal of misinformation content, could more generally affect any of the above subclasses.

4 Platform Policy and Platform Rules

Finally, the combination of existing ontologies and classes must lead to the facilitation of user-managed and co-informed misinformation handling. These protocols are encapsulated by **"rules"** such as:

Rule
I want to be notified of any possible fake news about refugee migration

that adhere to previously agreed-upon **"policy"** such as:

Policy
Potential fake news will be reviewed by fact-checking and then revised or validated

In this example we see the relationship between rules and policy, as defined in Section 1.1.2: the policy sets up general directives for the platform's misinformation protocol, while the rules are then the implementations of this policy directive. On Co-Inform in particular, these rules are automated.

Thus, the user should leverage this feature by creating misinformation-related rules that are difficult to do by hand or even by a team of human staff. A visual representation is as follows:

Policy
Potential fake news will be reviewed by fact-checking and then revised or validated
<table><tr><td>**Rule**</td></tr><tr><td>▪ *I want to be notified of any possible fake news about refugee migration* ▪ *I want also to automate the following action…*</td></tr></table>

4.1 Co-creation of Platform Policies for Misinformation Management

It is impossible to know all required policies beforehand which applies to Co-Inform developers as well as platform users. Therefore, it is impossible to computationally represent from the beginning every thinkable policy. Because of the co-creation feature, users of Co-Inform actively engage in policy creation and policy update for management. They shall be able to search existing policies and discuss on policies with other users. In practice, this can be feasible by including existing mechanisms, such as a semantic wiki system, akin for example to Wikipedia, allowing for the continued maintenance of platform policies and user engagement [Butler, 2007]).

As specified in Section 1.1.1, policies are text describing guidelines and must be handled via code specifically written for a single policy (in Wikipedia this would be called a bot). In this manner, Co-Inform platform would be managed by community even after termination of the project, by maintaining elasticity and flexibility for evolving policies.

An initial core of policies is provided in Section 5, tackling different types of content allowed/restricted on the platform as well as some policies regulating the users' behaviour when engaging in co-creating the policies themselves.

4.2 Platform rules: Event-Condition-Action

A **Platform rule** can be presented as ECA (Event-Condition-Action) rule structure that defines that when an event is detected, the condition is evaluated, and if the condition is satisfied, it provokes the action. Formally, an ECA rule has the following core components:

- **Event:** describes the signal that triggers the invocation of a rule.
- **Condition:** specifies what should be checked to execute a rule.
- **Action:** specifies what to execute in response to the event.

In practice, ECA rules are "reactive" because a rules module reacts when something happens on the platform. Thus, in an ECA rule, an **Event** happens under a certain **Condition**, triggering an **Action**. The below Figure 11 illustrates this automated workflow[12]:
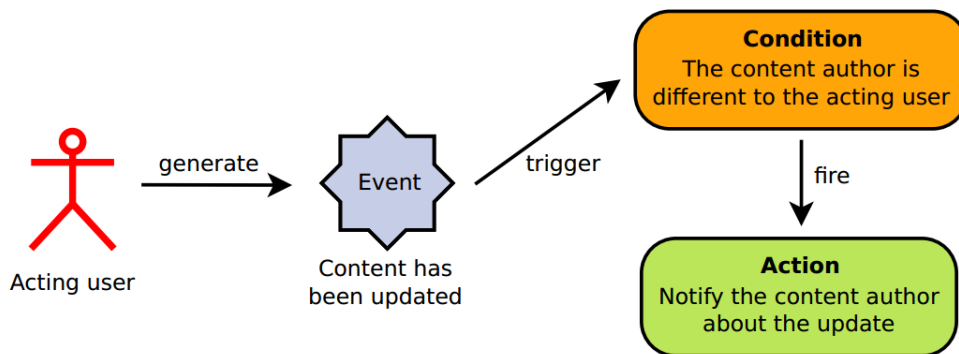
Figure 11. Single reactive ECA rule

As stated in previous sections throughout this deliverable, the Co-Inform platform aims to counter misinformation based on user needs that are a) identified prior to platform implementation (through interviews and surveys as in Section 1.2.1) as well as b) emergent

---

from actual usage. For the ECA rules, this means that the rule conditions will expand from their initial basic formulations to gradually cover and account for all cases of misinformation that users observe and wish to counter. In this sense, a hypothetical rule system that is more progressed and elaborated in comparison to Figure 11 is shown in Figure 12.
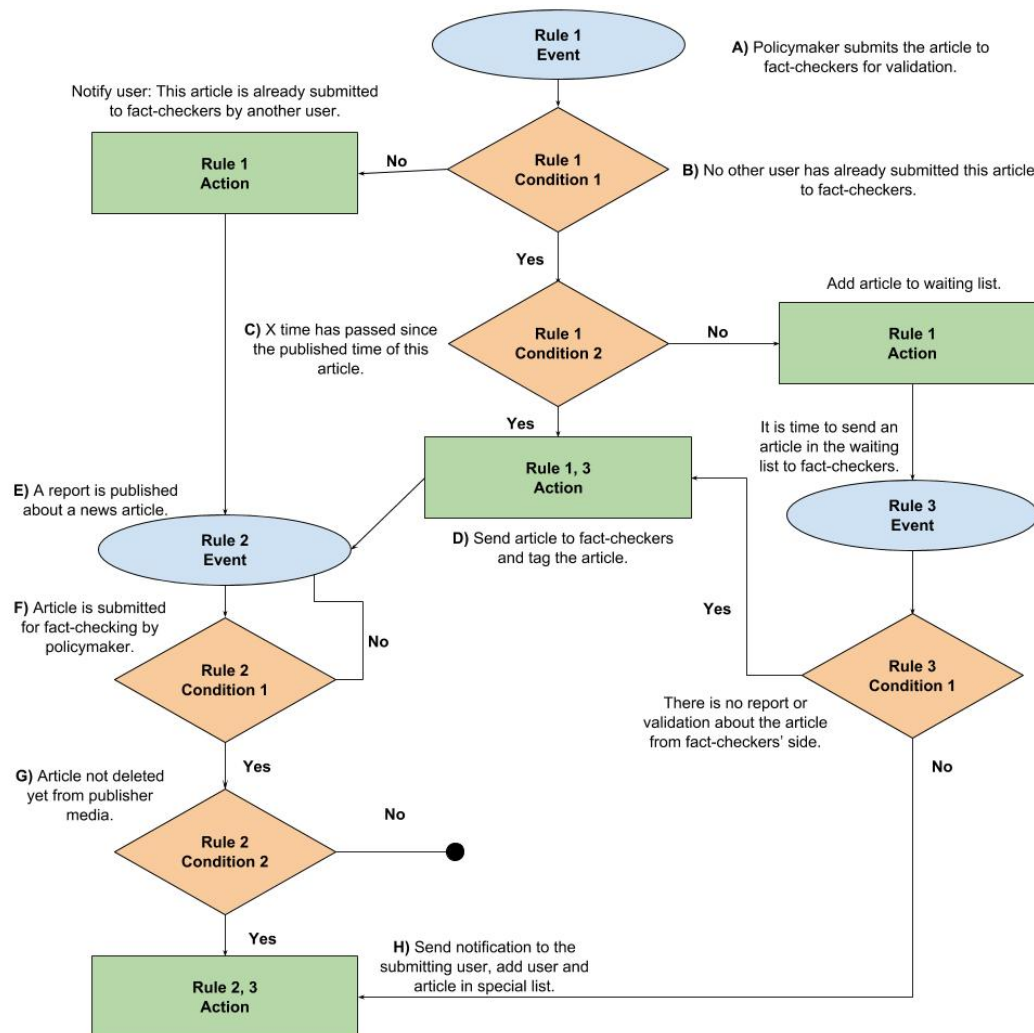

Figure 12. System of ECA rules in form of a scenario

The following is a hypothetical scenario that takes one route among possible options:

A. **Event I:** User 1 submits an article for validation.
B. **1st condition:** Has another user submitted this article prior? (If so, then User 1 will be notified that this article is already submitted)
C. **2nd condition:** If no other user has submitted the article prior, has enough time passed since the time of publication of the article?
D. **Action:** If enough time has passed since the time of publication of the article, it will be sent to fact-checkers who tag the article accordingly.

The above ECA chain can be followed by a second ECA chain as below:

- E. **Event II:** A report is published on a news article.
- F. **1st condition:** The article is submitted for fact-checking again.
- G. **2nd condition:** The article has not been deleted yet by the publisher media.
- H. **Action:** If the above two are the case, a notification is sent to the user. The user and the article are added to a special list.

4.2.1   Editing platform rules

Platform rules are in fact editable only because they can be actively edited, e.g. written by authors who are everyday users of the Co-Inform platform. The advantage of having reactive ECA rules is that site users can be provided with a powerful user interface to create custom automated workflows on a website, without any coding. The below Figure 13 illustrates the relationship of the author (Co-Inform user, who can be captured in terms of either sioc:UserAccount, foaf:OnlineAccount, or foaf:Agent) to editable platform rules, all of which follow the ECA grammar.
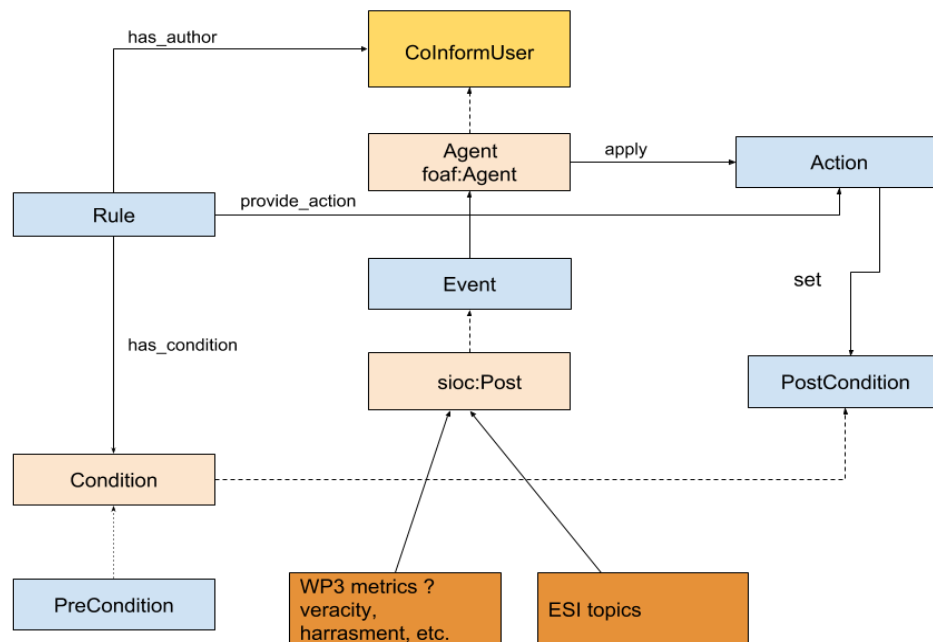

Figure 13. Formalizing user input into rules

4.3 Rule Automation Framework

We propose a framework utilizing semantic technologies to define and execute automated rules for misinformation management and interventions. The use of semantic technologies provides flexibility to implement rules and integration without change of source code of Co-inform module.

The Rule automation framework will be able to manage events coming from different sources and accordingly will trigger the proper actions generated by rule engine in the framework. And also, our proposed platform includes functions for creating and editing rules, as well as functions that handle automation of rules and the repository for storing rules in triple format. The ultimate version of the platform will receive events, evaluating them with rules in repository and executing the corresponding actions.
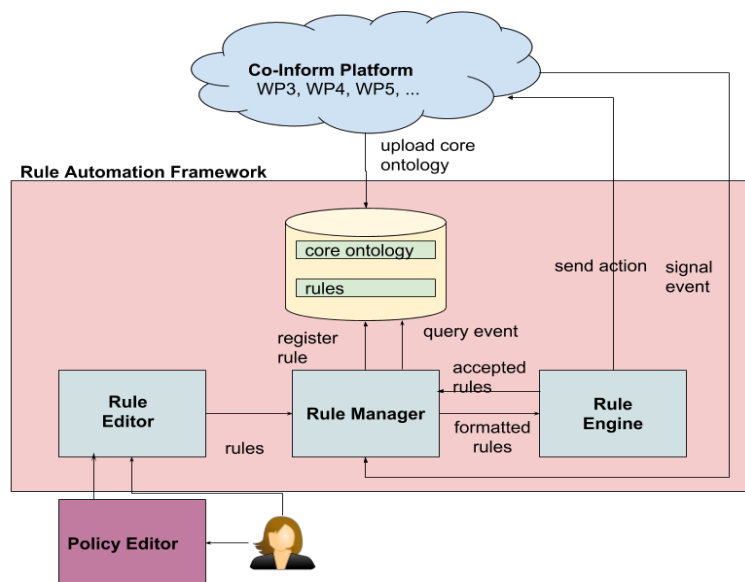

Figure 14. System Architecture of Co-Inform

The first draft of rule automation framework is depicted in Figure 14. Components of the rule automation framework are three-fold and they shall be used as reference and communication means for related Work Packages:

- Rule Editor
- Rule Manager
- Rule Engine.

In the following section, these three modules are described in detail.

4.3.1   Rule Editor

The rule editor is a user interface to create or modify rules. In the start window (Figure 15) Figure 15. Sample interface of Rule Editor (start window) a concise verbal description of what the rule should do, and then either adds, edits, or deletes the rule. The main window is shown in Figure 16.
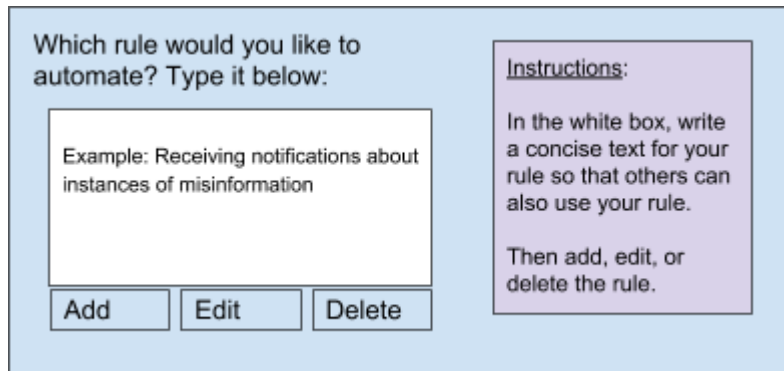

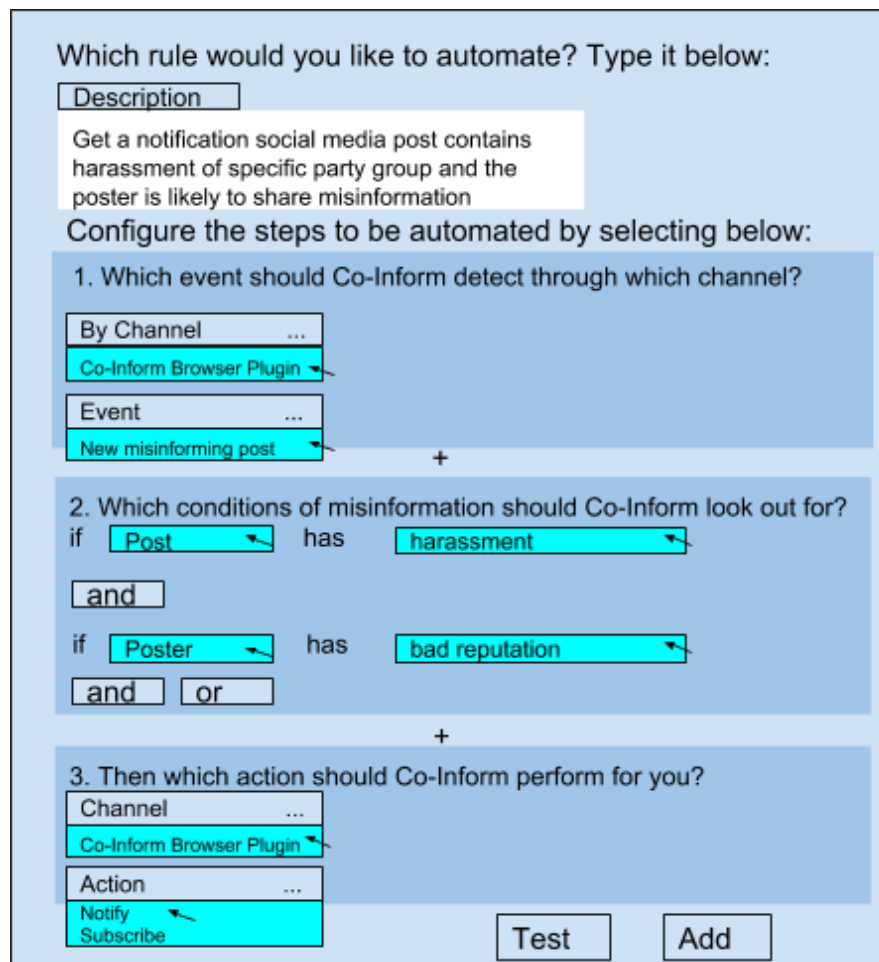**Figure 15. Sample interface of Rule Editor (start window)**

Figure 16. Usage of Rule Editor in a sample interface for Co-Inform Users

Buttons for adding and editing rules connect to popup window where users fill rule template as seen in Figure 16. The template uses essential components of Co-Inform ontology described in Section 3. When user configures component of the rule, rule manager retrieves entities corresponding component. For example, when user clicks Action button, policy editor sends the input to rule manager, and then rule manager queries all actions in the triplestore and sends the query results to the rule editor. Then user selects "Notify" action, his/her selection is appeared at the right side of the button.

When user configures all components, the input constructs a rule in JSON-LD format, which consists of several line of code[13]. The test button sends the code input to the Rule Manager. Then, the rule is saved by the rule manager to triplestore.

4.3.2 Rule Manager

The rule manager is responsible of CRUD operations for rules on triplestore. It also encodes the rule into format compatible with rule engine.

Rule Encoding

Rule automation framework utilizes third party rule engines and each rule engine supports different rule language. Rule manager accepts input rule as JSON format for interoperability between the rule languages and it encodes the format that is accepted by rule engine. First version of rule automation framework uses Jena rule reasoner and Jena rule language. The Jena rule language format is supported by Jena rules engine. The syntax of Jena rule is as follows:

[{rule label}:	List<condition>	->	List<consequences>

The syntax formulates "If List<condition>, then list<consequences> happen". Conditions and consequences are written as triples.

Example 4.1: Rule is defined by Harry

Get a notification if social media post contains harassment of specific party group and the poster is likely to share misinformation

Rule manager converts the rule defined in Example 4.1 as shown in **Example 4.2.**

Example 4.2: Jena rule format of **Example 4.1**

---

[13] The standard Co-Inform user does not need coding

[ruleNotify: (?post rdf:type sioc:Post) (?post flame:hateSpeech ?hscore) (equal(?hscore, true^^xs:boolean)) (?post sioc:sharedBy ?user) (?user authoritativeness:PosterReputation ?rScore) (lessThan(?rScore,"some_value")) -> (?coinformUser coinform:name "Harry") (?action rule:assigns ?channel) (?channelName channel:name "Co-inform browser plugin"^^xsd:string) (?channel rule:target ?coinformUser) (?action rule:action "notify")]

Rule CRUD

- **Creating new rules**:

Once rule is mapped to compatible format by rule manager, rule manager sends the rules to the rule engine which checks the rule whether it is compliant with the core ontology. Accepted rule is registered to triplestore.

- **Deleting existing rules:**

Rule manager send a delete request to remove the rule from triplestore.

- **Update existing rules:**

Rule manager send an update request to update the rule in triplestore.

Event Listener:

When an event occurs, rule manager is triggered by Co-Inform platform. Rule manager queries set of rules corresponding the event and send the query results to the rule engine. Rule engine executes the rules and then send the response to the platform.

4.3.3 Rule Engine

Rule engine is core component of rule automation framework. It acts as filter mechanism and is responsible of resulting in corresponding actions when an event is triggered. It requires core ontology described in Section 3 and rules for verification of new rule or execution of existing rules in triple store.

As rule engine, generic rule reasoner of Jena will be used. Generic rule reasoner supports external rules. It can be configured as forward chaining, backward chaining or hybrid engine[14].

---

[14] https://jena.apache.org/documentation/inference/

5 Initial Platform Policies

In this section, we bootstrap an initial battery of platform policies. In this endeavour we drawn inspiration from the results of the Citizen Focus Group as well as from existing platforms like Reddit, Wikipedia, Twitter and Facebook.

Of course, the co-creation nature of the **Co-Inform** platform implies that this initial set is just the seed, and, through the co-creation process, the users are endowed with the capacity of practically extending and modifying it. For the sake of clarity, we made a first separation between policies regarding content and the ones focusing on the users, which we detail in the two below subsections.

5.1 Content Policies

These policies describe what is and what is not possible to publish on the **Co-Inform** platform. In what follows the word "content" designs both Posts, which have the dedicated sioc:post class, and users' Comments, which have the schema:comment class of the **Co-Inform** ontology (Section 3.5).

5.1.1 Posts Policy

Posts on the **Co-Inform** platform have the latitude to contain potentially sensitive material and language register as an expression of misinformation (see Section 3.6). However, they are comprised in the Restricted Content Policy (Section 5.1.2).

5.1.2 Restricted Content

Users ought to refrain from publishing content that:

- Is illegal
- Is pornographic
- Involves minors in sexual and/or suggestive manner
- Is spam
- Contains personal and confidential information

5.1.3 Guidelines for Users' Comments

The Co-creation process engages the users to modify, add, remove policies and rules, through the Policy Editor and the Rule Automation Framework (Section 4) in order to better suit their needs.

Co-Inform users can discuss modifications to **Co-Inform** Policies through the Policy Editor, however such discussions must be led in a non-offensive manner.

Therefore, in addition to the restricted content, **Co-Inform** users' comments must not contain:

- Hate speech
- Encouragement to violence
- Threats and harassment

5.2 User Policies

On the **Co-Inform** platform, users can submit content, which is potentially misinforming, to be fact-checked.

To this end, it will be reviewed and, then, a report shall be published on the platform. Such report will be available to Co-Inform users to access and to share on social media, which is encouraged.

Co-Inform users are also encouraged to actively maintain the discussion around policies civil and meaningful, thus flagging comments which do not align with **Co-Inform** Content Policies.

In case of infringement of the Content Policies, a three steps intervention is envisioned according to the offence gravity and recidivism:

- Warning to the offender
- Severe warning and temporary suspension of the account.
- Deletion of the offender's account

6 Appendix

6.1 Semantic Web Technologies

We aim to implement the prototype of WP2 framework by using semantic web technologies.

Apache Jena: is an open source Java based framework which provides a programmatic environment for Semantic Web Technologies and includes rule-based inference engine. We will utilize Java library of Apache Jena to implement core functions of semantic engine where are used in manipulation and retrieval of RDF data over triplestores. Additionally, module of Jena for rule inference will be integrated in rule engine of the framework.

SPARQL: is a query language and a protocol which is used to extract information from RDF data. Apache Jena will be used to create and to execute SPARQL queries.

Apache Jena Fuseki is sub-project of Jena and is a SPARQL server. Fuseki will be used as storage of domain data and rules of WP2 and also endpoint to query the knowledge graphs. A screenshot of how querying RDF graph in Fuseki Server is shown in Figure 17.

JSON: is a lightweight text format for exchanging data. We will express initially commands as output of rule engine in JSON[15] format.

Protege: is a toolkit for ontology development. Co-Inform domain ontology and ECA ontology will be developed by using Protege. Protege will also be used to test initial rules which will be executed later in the framework.

---

[15] **https://www.json.org/ retrieved on 15.01.2019**

Figure 17. Apache Jena Fuseki GUI

6.2 Results from Co-Creation Workshops

In this section, we present first results of co-creation sessions held in Sweden and Greece.

6.2.1 Co-creation workshop in Sweden

Results from First Co-creation workshop held on February 15th, 2019 in Botkyrka, Sweden that can inform ECA platform policy rules are given in Table 11.

Table 11. Platform Policy Rules		
Event	**Condition**	**Action**
Did Sweden Rename Christmas to Winter Celebration?	Target audience - people outside of Sweden. As Sweden has become the symbol of a liberal / leftist society gone wrong, within this context. Audiences in Sweden know this is not true. False portrayal of Sweden as being taken over by a multi-cultural left-wing conspiracy.	Don't legitimize or make credible via repetition/ amplification. Contextualize this information - probably based on true events. Also, in other countries, referred to as "seasonal greetings". Also present specific instances of where this occurred, and did it cause any conflict or real crisis? Find out why people feel the need to spread this? What are the underlying tensions and fears? Provide a forum where this can be discussed to promote a shared understanding.
SAPO misinformed about a refugee who was arrested in 2015 as a terrorist only later to be released as innocent but with a damaged reputation for life. Media published photo and name and personal details of him without first confirming.	Specific migrant groups, police, municipality, citizens are all targets of this misinformation - this is done with the intent to spark fear and paranoia.	Public Database, where information can be verified. Active Citizenship concept brought up here - to allow the checking of the legitimacy of a source.
"No go Zones" - several separate events convene or	Stigmatize certain regions, associated with higher	Map/ database of what event happens where.

are stitched together in a way to portray geographic areas in Sweden where crime has risen to make them difficult to enter even by authorities.	diversity of residents, migrant populations and foreigners - to create the misinformation that they bring with them higher crime rates and aggression to authorities.	Positive articles about such communities to counter balance and provide another narrative, that is not all negative - disproportionally so. Highlight who are the victims of this narrative - give them voice and show who they are. Are algorithms at play here to create bias against certain communities - provide greater transparency for this. Browser plug in for this - if you type in No Go Zones - it provides critical, positive articles to balance the negative rhetoric. Also provides a meta level analysis on what you're reading and why?

Additional Notes:
- There is a need for Agency as opposed to Reactive mitigation of misinformation - we need our own narrative not just feeding into the rhetoric of misinformation.
- We need to deconstruct images and assumptions behind images. We can counter misinformation of negative messages via positive ones.
- We need to use the language of the gut, emotions, not just Black and White facts to counter misinformation - for it to stick.
- We need a commons approach to ownership of media = i.e. if its privately owned then the legitimacy of source and trust will always be compromised.
- Finally, time frame is key - when policy makers/ journalists are too rushed to check facts, errors occur.

6.2.2 Co-creation workshop in Greece

In the Greek Pilot workshop, a joint session of the stakeholder was held on March 21st, 2019, where we collected feedback on the policy scenarios from the stakeholders (Policy makers, Citizens and Journalists). The moderator presented the scenarios shown in Table 12. All three focused group deliberated these misinformation management polices scenarios and proposed the actions as a feedback which is helpful for platform policy rules. The feedbacks from stakeholders in terms of action are detailed in Table 13.

Table 12. Policy Scenarios presented in Greek Pilot Workshop			
Description	**Objective**	**Event**	**Condition**
Scenario 1			
Suppose there is news published in the social media which contains fake fact information and very harmful to the whole nation. e.g. there is dispute between political government and armed forces of the country. So, there is possibility of happening like last time done in Turkey.	The objective of this exercise is to determine the specific actions of the key stakeholders about fake news based on their existing approaches in the Greece.	Disinformation news published in media (conventional and online social media)	The news contains some previous facts which are not current. However, it creates confusion for the recipients of the news to determine its creditability.
Scenario 2			
There is news about shortage of beans in the county due to some change environment as per the analysis of some experts' global environmental changes. There is big chance that in next couple of months we need to import beans from other countries and our people will get the beans may be 4 to 5 times higher prices in the market.	The objective of this exercise it to determine the mechanism of trust the people of Greece opted in case of news	Disinformation is published in online social media (newspaper) having access of this web media source is about 5 thousand users (user count per day)	The news does not describe the exact source of the information and other news media did not cover this news

Scenario 3			
Every day, there are several news in the social media which are published based on the intent of the authors i.e. to gain profit, for ideological/political purposes, or intentionally cause public harm. The creation and dissemination on a large scale and with high speed that is unprecedented. In view of the above, do you think that there should be technological platform known as "fact-checking tools" to deal with a huge number of misinformation /fake news	The objective of this exercise it to determine their viewpoints on fact checking tools	A huge number of misinformation is created and disseminated on the social media	There is very less chance for a human being to deal with such disinformation manual

Table 13. Stakeholder Feedback in terms of Potential Actions		
Policymakers	**Citizens**	**Journalists**
Scenario 1		
i. I have to make a formal rebuttal with a press release" and send to reporters. ii. The press release comes to confront the reputation that has been created in society not only to act as an official source of information and for a means	i. I would watch the social media and see the first tweets on the subject from which means and persons they come from. ii. I will first look to see if this news has risen to the most reliable media for me and then I will contact people in my circle who I trust to discuss why everyone has different filters.	i. I would be able to contact some sources that could give me information so that I can immediately post what is happening. ii. However, the rumour may have taken importance's by then Dimensions. "when propaganda flows as much as I do, the fame will be established as a reality. "

Scenario 2		
i. I will look at the source and for me in this case it is the ministry. ii. Missing the piece that the citizen has knowledge and potions to check reliability. I don't think any news Check from a search engine from any person (online). Important to see the social structure of those who accept and believe any false news iii. Missing the piece that the citizen has knowledge and potions to check reliability. I don't think any news Check from a search engine from any person (online). Important to see the social structure of those who accept and believe any false news.	i. I would like, as a citizen, that journalists' control who these specific experts are, if they are known and verified and what is their motives? So, then I may know if I'll trust the News. ii. We'll all go to the supermarket and get supplies iii. For any information we will not be able to be 24 hours in front of a computer to control reliability. Ignore that every person has knowledge in specific fields. I'm going to ignore the news' cause I know iv. We'll all go to the supermarket and get supplies v. For any information we will not be able to be 24 hours in front of a computer to control reliability. Ignore that every person has knowledge in specific fields. I'm going to ignore the news' cause I know	I'll try to get the original source from those accounts that spread out the news
Scenario 3		
It is Important that the journalist/citizen can understand the value of the data and can analyse it correctly. For example, because there are many statistics that are susceptible to specific management.	i. Tools that give access to databases to cross-reference information or else a database of articles. ii. Groups that make online debate in articles, that would be a good idea. And as the journalist network grows, it 'll be even better. Also pass this fact checking very fast in Society, the false news is spreading. iii. Global trend for subscription to trusted sites that do not accept payments from anywhere else except from	i. Education of people (including Journalists) on how to tackle misinformation is important. Methods, guidelines should be provided along with tools to use maybe. ii. It's Important to know who has the data. We usually ask for open data, but

	subscribers, the public (limiting the possibility of being directed by interests). iv. Groups that make online debate in articles, that would be a good idea. And as the journalist network grows, it 'll be even better. Also pass this fact checking very fast in Society, the false news is spreading. v. Global trend for subscription to trusted sites that do not accept payments from anywhere else except from subscribers, the public (limiting the possibility of being directed by interests).	the actors eventually do not have or give these data. They ask who you are, what do you intend to do with them, etc.

6.2.3 Co-creation workshop in Austria

In the Austrian Pilot workshop, a joint workshop with stakeholders from Austrian Limited Profit Housing Sector as well as policy-makers and journalists who deal with questions of migration and misinformation was held on March 28th, 2019. During this workshop we collected perceptions and views from three major groups of stakeholders: policy makers, citizens and journalists (Table 14).

Table 14: Perceptions of events and sources of misinformation as well as of conditions for implementation of tools and methods to deal with misinformation and recommendations for policy actions among policy-makers, journalists and citizen

	Event: **From which sources do you receive information about events important for migration issues?**	**Condition:** **Which conditions are necessary for implementation of methods and tools to fight misinformation in the daily praxis?**	**Action:** **Recommendations on actions to fight misinformation**
Policy-makers	Internet (100%) Newspapers (71%) Personal talks (friends, colleagues, family) (57%) Email (43%) TV and radio (29%) Statistics (14%)	Existence of several information sources (no monopoly of one media). Wish of people to read, check information, to be critical, to search for alternative sources. Culture of discourse and possibility for different discourses to be open in public space.	Further spread of Wikipedia and alternative information sources in Internet. Spread of printed traditional media online. Creation of info-points and service points to provide information about migration. Sufficient opportunities for networking and personal discussions about issues relevant for migration. To provide fact checks in traditional

			media such as ORF online. Development of regulations for social media regarding misinformation. Creation of conditions which guarantee independence of media and research institutions. Introduction of institution of online journalism
Journalists	Social media (100%) Newsletter (86%) Twitter (57%) Podcasts (43%) Personal talks (29%)	Awareness about existence of misinformation. Awareness about credibility of different information sources and avoidance of sources which provide misinformation. Possibilities to cross check twitter accounts and to verify their sources.	Measures to strengthen awareness that behind every news could be a special political agenda. Blacklist of information sources which provide misinformation. Critical thinking about from whom and why this information is coming. Usage of plug-ins in browser. Cross checking by fact-checking sites. Implementation of events of awareness raising about misinformation and media literacy.

			Implementation of events on sensibilization in social media about misinformation. Providing easily accessible tools which show sources of information online
Citizen	Newspaper (100%) Internet (86%) Facebook (57%) Personal talks (43%) Books (29%) Events (14%)	Existence of culture of critical thinking and willingness to search for alternative information. Possibility to use several sources of information. Personal experience and education.	Provide recommendations in traditional media about reliable sources of information. Organize public events about reliable sources of information and to raise awareness about misinformation. Publish media reports about "from whom comes bad information". Create culture of thinking when people don't hurry up with conclusions but are searching for alternative information.

References

[Butler, 2007] Butler, Brian, Elisabeth Joyce, and Jacqueline Pike. "Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in Wikipedia." In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1101-1110. ACM, 2008.

[Noy, 2001] Noy, Natalya F., and Deborah L. McGuinness. "Ontology development 101: A guide to creating your first ontology." (2001).

[Fernandez, 2014] Fernandez, M., Scharl, A., Bontcheva, K., & Alani, H. (2014). User profile modelling in online communities.