EU COMMUNITY

ICT-2013.5.4 ICT for Governance and Policy Modelling

EU COMMUNITY MERGES ICT AND SOCIAL MEDIA NETWORKING WITH ESTABLISHED ONLINE MEDIA AND STAKEHOLDER GROUPS TO CULTIVATE TRANSPARENCY, ENHANCE EFFICIENCY AND STIMULATE FRESH IDEAS FOR EU POLICY-MAKING

Deliverable D3.2.3

Reputation Management Module

(Software/ Documentation)

(Third version)

Editor(s):	Manolis MARAGOUDAKIS, Miltiadis KOKKONIDIS, Aggeliki ANDROUTSOPOULOU, Demetrios PYRENIS
EU COMMUNITY:	UNIVERSITY OF THE AEGEAN – RESEARCH UNIT
Status-Version:	Final – v1.0
Date:	04.09.2016
EC Distribution:	PU

Project Number:	611964
Project Title:	EU COMMUNITY

Title of Deliverable:	Reputation Management Module (Software/ Documentation) (third version)
Date of Delivery to the EC:	05/09/2016

Workpackage responsible for the Deliverable:	WP3 – Opinion Mining and Reputation Management Component
Editor(s):	Manolis MARAGOUDAKIS, Miltiadis KOKKONIDIS, Aggeliki ANDROUTSOPOULOU, Demetrios PYRENIS
Contributor(s):	AEGEAN, INTRA, UOC
Reviewer(s):	UOC, EURACTIV
Approved by:	All Partners

Abstract:	The key task of the Reputation Management Component is to assign a reputation score to each person of interest per topic. This reputation score is a synthetic score computed as a combination of scores for a number of reputation criteria. The Reputation Management Module in its third version implements all the criteria specified in D2.4 (Community Requirements and Specifications) as well as additional ones. Synthetic reputation scoring also needs a way of combining these individual scores. Both the method envisaged in D2.4 and additional ones are supported in the Reputation Management Module. In accordance with the delivery plan, work on

	the third version of the RMS was meant to focus on testing, debugging and solidifying the work already seen in the second prototype and become the intermediate evolutionary step from a series of prototypes to being part of a commercial product or online service. Its maturity was demonstrated in a high-profile event in the European Parliament on the 24th of February 2016, where the first Consortium-validated Energy Union ranking was unveiled. However, in order to satisfy additional requirements that emerged out of WP4, work on the RMS continued, resulting in an second release with an additional criterion, computed in collaboration with the Hybrid Predictions Subsystem (HPS), rewarding experts that make correct predictions on the outcome and duration of legislative procedures. The third prototype also addresses concerns arising from the fact that the RMS was designed from the outset to provide topic-specific scores and rankings, as it describes a way that topic-specific scores can be combined in order to compute an overall topic-independent score. The present deliverable describes the effort put into validating the results of the Reputation Management Module and the methodology for adjusting the weights of the various criteria. It also addresses issues related to private vs. public data and explains the decision to only make public the ranking of experts and a general description of the individual criteria.
Keyword List:	Reputation Management, Synthetic Reputation Score, policy process, Crawlers Component, Europa WhoIsWho, Twitter, LinkedIn, social graph.

Document Description

Document Revision History

Version	*Date*	*Modifications Introduced*	
		Modification Reason	*Modified by*
0.1	15/01/2016	Initial draft version on the basis of D3.2.2. Circulated to INTRA and HSE for comments and contributions	AEGEAN
0.2	04/02/2016	Revision of implementation and chapter explaining how the RMS produces a generic (i.e. non topic-specific) reputation score	AEGEAN, INTRA, HSE
0.3	29/02/2016	Inclusion of Euractory Top 40 Ranking Results for Energy Union presented at the European Parliament on February 24, 2016	AEGEAN
0.4	30/02/2016	Updating of Integration section	HSE
0.5	03/03/2016	Addition of first draft of Adjustment of Weights Methodology section	INTRA
0.6	04/03/2016	Addition of first draft of Experts Feedback section	AEGEAN
0.7	17/04/2016	Addition of Predictions criterion and revision of the deliverable where necessary	INTRA
0.8	18/04/2016	Revision and re-organisation of entire deliverable	AEGEAN
0.8a	21/04/2016	Minor edits	HSE
0.8b	03/05/2016	Minor revision of Expert Input Collection Methodology and Results. Minor edits.	INTRA
0.8c	04/05/2016	Minor edits. Version submitted for review.	AEGEAN

V0.9	01/09/2016	Finalised details of deliverable on the basis of external experts' comments and discussions with WP4, WP7 and WP6 leaders focusing on the Predictions Score criterion. Final (minor) edits.	AEGEAN, INTRA, HSE
V1.0	02/09/2016	Final version to be submitted.	INTRA

Contents

List of Figures

List of Tables

Definitions, Acronyms and Abbreviations

Table 1: Acronyms and Abbreviations

Acronym	Title
JSON	JavaScript Object Notation
RMS	Reputation Management System
SQL	Structured Query Language

Table 2: Definitions

Acronym	Title
Uni-Dimensional RMS	A Reputation Management System that provides a single reputation criterion
Multi-Dimensional RMS	A Reputation Management System that provides multiple reputation criteria
Synthetic Reputation Criterion	A Reputation Criterion whose value is a function of the values of more primitive reputation criteria and a number of configuration options
Synthesis-Enabled RMS	A Reputation Management System that includes one or more Synthetic Reputation Criteria
Synthetic RMS	A Uni-Dimensional Reputation Management System that provides a single synthetic reputation criterion (computed in terms of more primitive reputation criteria which are only used internally and not exposed)
Objective Reputation Criterion	A reputation criterion that has the same value irrespective of the observer i.e. a criterion that will have the same value for all users.
Subjective Reputation Criterion	A reputation criterion the value of which depends on factors including who the observer is i.e. a user-sensitive criterion.
Personalisation (in the context of RMSs)	The ability of an RMS to support subjective criteria and/or per-user configurability of (some of) its options.

1 Executive Summary

This deliverable outlines the design of the Reputation Management System of the EU Community platform as it currently stands (version 3).

The EU Community RMS combines a number of interesting characteristics:

- It is a synthetic reputation scoring system. Synthetic reputation scoring is typically used in cases where a single actionable score is needed, either for providing a ranking (e.g. web search results) or making a decision (e.g. should a loan be issued to a particular bank customer or not). In EU Community, synthetic reputation scoring is used for both. The computed synthetic reputation score is the result of synthesis of a number of individual `reputation' scores. Whereas the individual scores could be used directly for the same purposes, the resulting decision trees or sorting criteria would either be too simplistic (and fail to capture the same information the synthetic score does), or too complex (resorting to ad hoc reputation synthesis operations, which, typically, both increase complexity and reduce transparency). The synthetic reputation score computed by the RMS provides a convenient unified measure of reputation per person per topic which makes both decision-making and ranking much easier for the various EU Community components than a multitude of atomic reputation scores would.

- The synthetic reputation score is the result of synthesis of individual reputation scores each based on a different reputation criterion and reputation source. These include self-assessment and peer assessment criteria as well as document-related and network-graph criteria. Reputation sources include the EU Platform itself as well as crawled sources.

- A synthesis function is responsible for combining the individual reputation scores for a given person and a given topic into a single synthetic score for that person-topic combination. The synthesis function specified in D2.4, as part of the requirements gathering process, was a simple weighted average function with fixed weights. The Reputation Management Component moves beyond that baseline specification, allowing users not only to specify custom weights for the available reputation criteria, but also to choose among a number of available synthesis functions as well. Administrators can make such choices for the default synthesis function.

- In order to allow such level of user-customisability, while providing real-time feedback on the effects of the user's options and supporting large-datasets, a separation of the processing stages and their corresponding input and output is necessary. Thus, a significant part of the computational workload is carried out off-line leaving behind data that can then be used for effectuating a user's choices in real-time.

- Interestingly, EU Community's reputation scoring is subjective, meaning that for different users, the score for the same person-topic combination may be different. This is due to the inclusion of subjective criteria which reflect the fact that a user is likely to attach more significance to his/her own network connections, document evaluations and so on and so forth.

Three versions of the Reputation Management System were scheduled for delivery. The present document describes the functionality offered by the third and final version.

2 Introduction

2.1 Objectives and Purpose

This report, entitled 'Reputation Management Module (third version)' is the third of a series of three deliverables of **EU Community** Work Package 3 (WP3). The purpose of each report in the series (D3.2.1, D3.2.2, D3.2.3) is to outline the design of the corresponding version of the Reputation Management System.

This deliverable is to fulfil the following objectives:

- Provide general background on the topic of Reputation Management.
- Present and classify the various pre-defined reputation scoring criteria.
- Present the various pre-defined synthesis functions.
- To present the user and administration customisability capabilities of the RMS.
- To present the algorithms of the third-version of the RMS.

2.2 Relation to other Work Packages/Deliverables

This deliverable is the direct output of Task 3.2 'Reputation Management Design & Component', one of the main tasks of WP3.

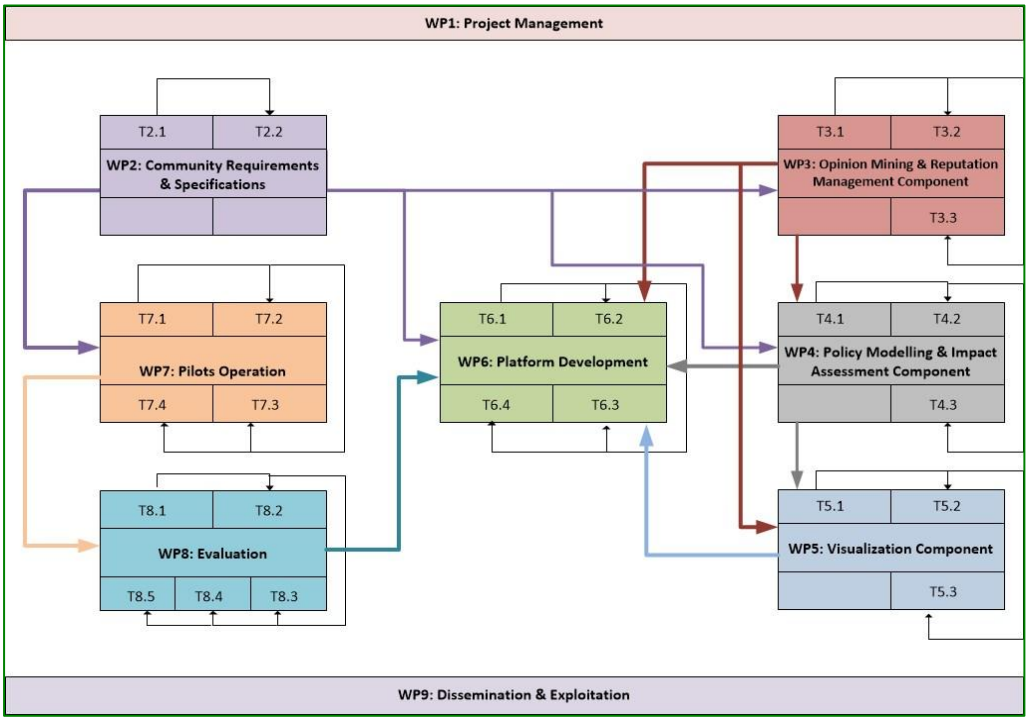

Figure 1: WPs Interdependencies and relations

The output of the Reputation Management Component is used by the user-facing components of the EU Community platform, namely EurActory (WP6) and PolicyLine (WP5). In turn, these components are relied on to provide input to the Reputation Management System. The RMS works by combining both user-input and publically available web content collected by the EU Community platform.

As illustrated in Figure 2:

- EurActory and the Crawlers Component provide data to CurActory (the former by recording user input, the later by importing external content)
- The Opinion Mining Component adds information to CurActory such as document sentiment information.
- The Reputation Management Component pre-computes and stores processed reputation data in CurActory; this is done off-line on a daily basis.
- Whenever reputation scores using custom user-selected synthesis functions are requested, the Reputation Management Component computes on-demand, in real time a synthetic reputation score on the basis of the users' selections and the pre-processed reputation data in CurActory.

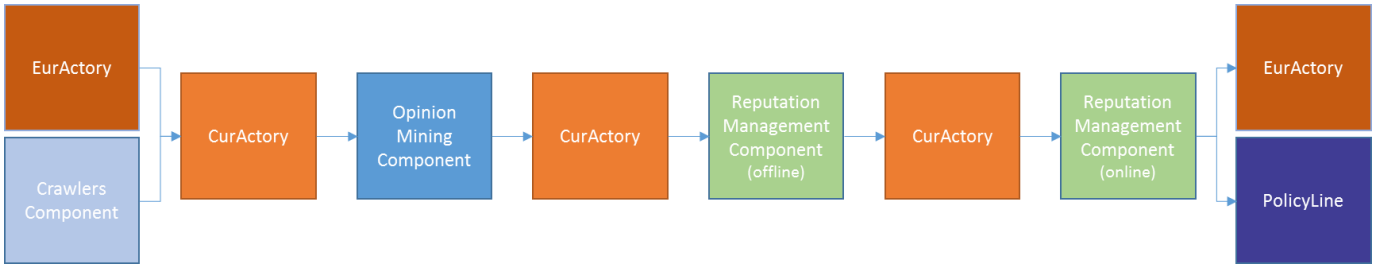

Figure 2: Component Interdependencies and relations

Initial requirements for the Reputation Management System were outlined in D2.4 Community Requirements and Specifications – Research Strategy. The previous (D3.2.1, D3.2.2) and current (D3.2.3) deliverables focusing on the EU Community Reputation Management System represent the Consortium's evolving understanding of how to best approach the complex subject of reputation management in the context of the EU Community project.

2.3 Structure of the Document

The structure of the deliverable is the following:

Chapter 1 provides an executive summary of the deliverable.

Chapter 2 provides a general introduction to the deliverable placing it in the context of the Work Package and the Project.

Chapter 3 provides a brief outline of the requirements set out for the EU Community Reputation Management system and a comparison with three noteworthy, for different reasons, extant reputation management systems on the web.

Chapter 4 outlines the Reputation Management System's architecture and its advantages over an initially considered alternative.

Chapter 5 specifies the reputation criteria found in the third version of the EU Community Reputation Management System.

Chapter 6 specifies the synthesis functions found in the third version of the EU Community Reputation Management System.

Chapter 7 specifies the behaviour of the topic- and observer-sensitive reputation scoring function when one of these two important parameters, topic and observer (i.e. current user) is missing.

Chapter 8 specifies the interfaces the EU Community Reputation Management System provides for integration purposes.

Chapter 9 explains the current methodology for adjusting the weights of the individual reputation criteria, explaining how its dynamic nature serves the aims of the project.

Chapter 10 presents feedback provided by WP3 External Experts on the design of the second version of the RMS.

Chapter 11 presents the methodology for engaging policy domain experts aiming at collecting the initial human expert-sourced policy domain reputation data.

Chapter 12 addresses the issue of privacy and presents the different classes of data used in the RMS.

Chapter 13 provides the conclusions and next steps for the RMS.

3 Background and Terminology

3.1 Approaching Reputation

The first and possibly most difficult concept to define in the context of a document describing a Reputation Management System is that of `reputation' as it is not a well-defined concept, resulting in the term being used to mean relatively different things.

Still, it is possible to attempt to provide a framework for how reputation is obtained. According to a plausible framework for reputation, an individual's reputation depends on:

1. Volume of work
2. Visibility of work (direct/indirect)
3. Quality of work

For instance, Dostoyevsky has a reputation as one of the great authors of all time due to having authored many famous well-received works. Had he authored a single great piece of literature, its reputation would overshadow his. Had his work not been published and read by the masses (direct visibility) and widely talked about (indirect visibility), he would not have had the reputation he currently has. Had his work been mediocre, but visible for other reasons, he could perhaps have achieved fame/infamy but not a reputation as a great author.

One shortcoming of the above reputation breakdown framework is that it fails to quantify reputation and the effect of the three proposed reputation constituents.

Another obvious shortcoming of this framework is that, though it is widely applicable, it does not cover cases where reputation stems from position (head of state, judge, CEO of Microsoft etc.) or other factors. In investigating what reputation is, one needs to keep a very open mind, adjusting any theoretical framework to the needs of the analysis rather than trying to fit the data into the framework.

In the context of Reputation Management Systems such as the one described herein, such theory/data mismatch issues do not / cannot exist: reputation is whatever the Reputation Management System says it is. In other words, each RMS defines its own sense of reputation. The question then is whether this is a *useful* notion in the context that motivated the existence of the RMS in the first place. So, an RMS can only be evaluated with respect to how well suited it is for achieving a specific *purpose*. Moreover, reputation metrics are quantitative by nature and if there is a synthesis of one or more factors this is clearly specified (although the algorithm is not necessarily visible to end users). The following section gives examples of RMSs in web applications.

3.2 Three Example Reputation Management Systems on the Web

Before presenting the reputation management system of EU Community, we examine three interestingly different reputation management systems, for comparison purposes.

3.2.1 The eBay RMS

One of the key game-changing features of the web, as it has evolved during the past few decades, is its ability to empower users to perform tasks where they are, as opposed to requiring their physical presence at specific points-of-service. With that came an opportunity to cater not only to multiple customers but also multiple suppliers, making sites such as eBay, Amazon, and Alibaba trading hubs, whereby hitherto unknown to each other buyers and sellers enter into business agreements.

The first Reputation Management System to be examined here is that of eBay because it has been and remains emblematic of e-commerce hub RMSs, because eBay is such a big player in the internet-enabled sales arena and because the eBay RMS is such an important part of what is eBay today for its users (buyers and sellers alike).

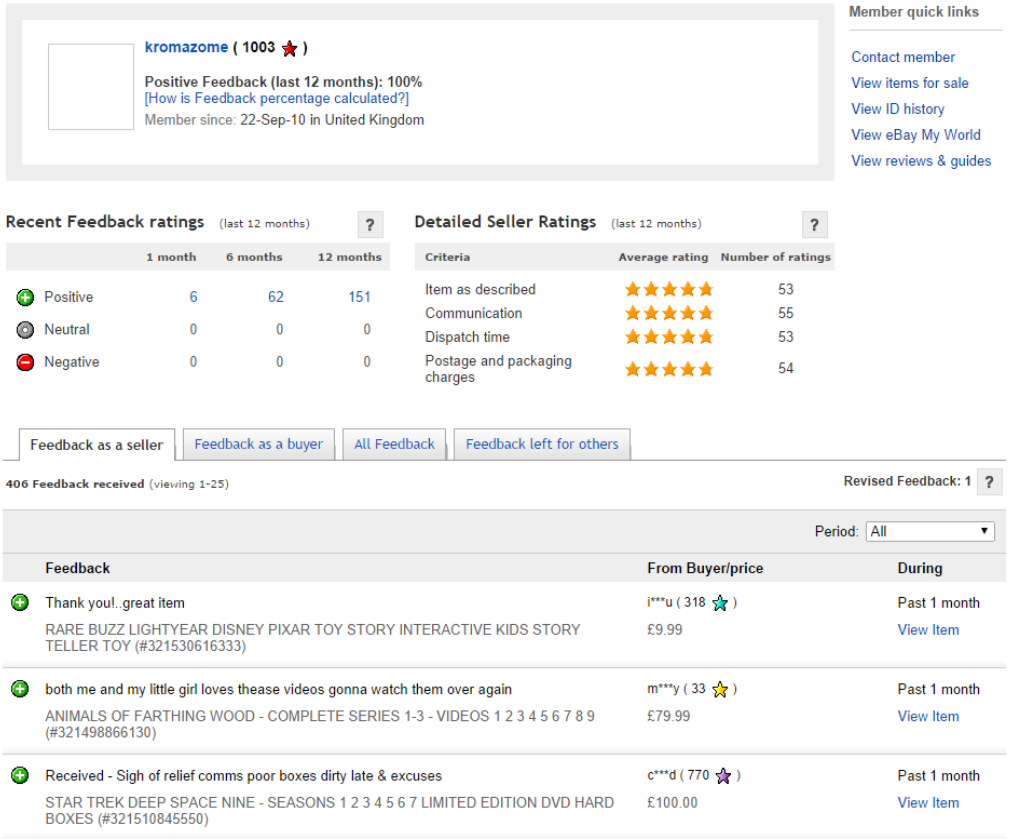

Figure 3: eBay Member Feedback and Reputation Page

The Reputation Management System in eBay supports (see Figure 3):

1. A cumulative feedback score: The number of positive (+1), negative (-1), and neutral (0) feedback ratings a member has received over time.

2. An optional star icon: Stars are badges of honour awarded on the basis of the cumulative feedback score (see Figure 4).

    a. A positive feedback, a negative feedback and a neutral feedback counter for feedback received during the past 1 month.

3. A positive feedback, a negative feedback and a neutral feedback counter for feedback received during the past 6 months.

4. A positive feedback, a negative feedback and a neutral feedback counter for feedback received during the past 12 months.

5. A percentage of positive feedback computed as the ratio of the above positive feedback counter value divided by the sum of the above positive and negative feedback counters (i.e. the percentage of positive feedbacks considers only feedback obtained during the past 12 months).

6. Short feedback messages accompanying (and explaining) the positive, neutral, or negative feedback:

7. Detailed seller ratings: A feature not originally found in eBay, detailed seller rating include 0 to 5 star ratings (computed as averages of relevant feedback received during the past 12 months) on the following four categories:

    - Item as described
    - Communication
    - Dispatch time
    - Postage costs

It is interesting to note that items 1 through 5 in the list above correspond to a single feedback action involving a choice between positive, neutral and negative. Item 6 requires the user giving feedback to add a short explanation for the feedback left and Item 7, an afterthought in the design of the eBay RMS, requires the user to provide a five star rating for up to four aspects of a transaction (some may be deactivated if non-applicable; for example Postage costs does not apply when shipping is free).

Clearly, the eBay RMS relies on eBay users to both provide reputation information and to interpret that information, paying attention to how such information is organised (feedback for the past 1/6/12 months, feedback as a seller/as a buyer/all feedback/feedback left for others). Yet, despite the wealth of available raw reputation data available for perusal, it is the computed positive feedback score percentage together with the computed accumulative feedback score that are the most important tools for judging the reputation of sellers.

Interestingly though, both these computed scores are misleading in the case of non-professional sellers, as they include feedback they collected as buyers. It is somewhat surprising perhaps but eBay's Reputation Management System does not differentiate between sellers and buyers, as much as one might expect, promoting the idea of a good eBayer over that of a good buyer (or seller). What

the reputation system of eBay is geared to is creating a community (of eBayers, not just buyers or sellers); and it has to a large extent succeeded in doing that. The fact that the term "eBayer" is quite commonly used by eBay's users is a sign of its success in that respect. Moreover, it encourages buyers to become sellers thus fuelling eBay's growth as a platform for the auctioning of practically everything under the sun and the company's financial growth (as more sales lead to more income from commissions).

Star	Color	Number of ratings
	Yellow	10 to 49
	Blue	50 to 99
	Turquoise	100 to 499
	Purple	500 to 999
	Red	1,000 to 4,999
	Green	5,000 to 9,999
	Yellow shooting star	10,000 to 24,999
	Turquoise shooting star	25,000 to 49,999
	Purple shooting star	50,000 to 99,999
	Red shooting star	100,000 to 499,999
	Green shooting star	500,000 to 999,999
	Silver shooting star	1,000,000 or more (Wow!)

Figure 4: The eBay Reputation Star System

3.2.2 The Google Search URL RMS

Users of the Google search engine may not be faced with numerous metrics and may not be requested to provide any feedback. Yet they are provided with a service that relies very heavily on a Reputation Management System which ranks the results of a search query not only by relevance, but by a combination of relevance (how closely the query matches a document found at a certain URL and even how closely the URL itself matches the search query) and prominence (how important the URL and its content are deemed to be).

In many ways, the Google Search URL RMS is the exact opposite of the eBay RMS. Google requires no user input nor does it disclose the information it uses in

its algorithms which themselves are in turn complex and for the most part secret. Reputation data are collected primarily by means of Google's crawlers. There is no association with financial transactions, a characteristic of the eBay RMS that protects it from manipulation, and due to the high value of placing high on Google's results list the business of Search Engine Optimisation flourishes. Some SEO techniques are welcome (ensuring the textual content of webpages is easily extracted and appropriate navigation is provided) whereas others amount to efforts to maliciously trick Google into assigning a higher prominence (e.g. link spamming) and/or relevance (e.g. hidden text relevant to certain popular queries meant to be crawled and indexed, but not seen by to human visitors) for certain queries and Goggle' staff is in a constant state of alert for detecting and thwarting such, so-called, black hat SEO machinations. Moreover, Google supports so-called personalised searches by means of taking into consideration past search history[1].

This is how Google themselves describe how Google search ranks URLs (blurring the distinction between prominence and relevance to a query, and leaving room for personalisation):[2]

When a user enters a query, our machines search the index for matching pages and return the results we believe are the most relevant to the user. Relevancy is determined by over 200 factors, one of which is the PageRank for a given page. PageRank is the measure of the importance of a page based on the incoming links from other pages. In simple terms, each link to a page on your site from another site adds to your site's PageRank. Not all links are equal: Google works hard to improve the user experience by identifying spam links and other practices that negatively impact search results. The best types of links are those that are given based on the quality of your content.

Unlike eBays's RMS for its members, Google's RMS for the URLs it crawls is not at all about transparency and community building, but rather about company trade secrets and ranking manipulation avoidance. PageRank may be patented andknown to the public, but this is only a piece of a large secret puzzle.

3.2.3 The Klout RMS

While Klout remains a small company, it has drawn considerable attention as they offer something relatively unique: a score for a user representing what they call 'influence', which in fact is a measure of how likely others are likely to react (within the context of social media) to that user's postings. They have an interesting business model whereby they provide their service to users for free and additionally arrange for top influencers to receive Perks from companies they work with. The idea is that these top influencers will then have nice things to say and that people will listen. Companies pay Klout for Perks campaigns, whereby they offer free services or products to select Klout users.

The company is not particularly forthcoming with details about the way the scores are computed. This is what they present as a "detailed description of how a Klout Score is determined":[3]

[1] https://support.google.com/websearch/answer/54068?hl=en Retrieved: 9-9-2014.

[2] https://support.google.com/webmasters/answer/70897?hl=en&ref_topic=4558960 Retrieved: 9-9-2014.

[3] https://klout.com/corp/score Retrieved: 9-9-2014.

The majority of the signals used to calculate the Klout Score are derived from combinations of attributes, such as the ratio of reactions you generate compared to the amount of content you share. For example, generating 100 retweets from 10 tweets will contribute more to your Score than generating 100 retweets from 1,000 tweets. We also consider factors such as how selective the people who interact with your content are. The more a person likes and retweets in a given day, the less each of those individual interactions contributes to another person's Score. Additionally, we value the engagement you drive from unique individuals. One-hundred retweets from 100 different people contribute more to your Score than do 100 retweets from a single person.

We know how important it is to maintain the integrity of the Klout Score, so we closely monitor activity across the signals we measure for inauthentic behaviors—spambots and the like. The Score will continue to evolve and improve as we add more networks and more signals.

Attempts to understand how Klout score works, for social networks such as Facebook, Twitter, Google+, Instagram, Foursquare, and LinkedIn suggest that what counts the most is how much others interact with one's postings rather than how many friends/followers/ etc. he/she has. This is consistent not with a system that attempts to estimate reputation in some generic sense but rather the likelihood than a Klout user will be an effective communicator of any positive opinions on a certain company that may happen to be offering Klout Perks. As mentioned earlier, the exact nature of reputation an RMS computes is very much dependent on the purpose for which the RMS was built and Klout's RMS has a specific business model to serve. However, in order to deal with criticisms about divergence from real-world reputation, as an afterthought, Wikipedia and Bing are now also used as reputation sources.

3.3 Key Technical Requirements

1. **Reputation Criteria:** After many discussions between the partners and with users during CreActiv workshops, the below criteria had been decided to evaluate the reputation and influence of EU stakeholders and formed the core of the first version of the RMS. Let O be the current user, A be an area of expertise and P be the person whose score is being computed for the area of expertise A.

    1. Self-evaluation: P's self-evaluation regarding his expertise in the area of expertise A.

    2. Organization Reputation: the reputation score for the organisation P works for.

    3. Position Reputation: the reputation score of P's job title.

    4. Document Assessment: a measure of the reputation of P's documents.

    5. Proximity Trust: a score corresponding to the level of connectedness of P with O.

    6. Network Value: a measure of P's social network's reputation.

    7. Past measurements: a measure of P's past reputation.

In the second version, due to reasons explained both D3.2.1 and in D3.2.2, the Organisation Reputation and Position Reputation criteria have been merged into a single criterion. Moreover, two new criteria were added in that version:

1. Peer Rating: a measure of the reputation P is assumed to have on the area of expertise A by other EU Community users.

2. Offline Reputation: a measure of P's real-world reputation (e.g. by considering both who is present in important events and asking those present who they think would have been a useful participant).

The third version of the RMS brings in a new criterion as a result of collaboration between WP3 and WP4. The new criterion examines how successful EU Community policy experts are at predicting the outcome and duration of legislative procedures.

2. **Weighted Average Synthesis Function:** Another requirement for the RMS is that the independent criteria scores be combined by means of taking their weighted average. All three RMS prototypes exceeded the relevant requirement by supporting a number of options for the synthesis function, including the weighted average synthesis function as the default.

3. **Configurability:** A further requirement was that the EU Community RMS must allow configurability, both by system administrators *and ordinary users* in the way it computes reputation scores, at the very least by allowing the weights of the various criteria to be configured (see Figure 5). This was a technical requirement meant to open up interesting possibilities if taken advantage of in EurActory. The configuration scenario for which the RMS was designed was that administrators will be able to configure the behaviour of the system reputation scoring, whereas ordinary users would be allowed to provide custom settings (which only apply to them).

4. **Performance:** It is a requirement that it should be possible for reputation scores to be computed instantly given a new set of configuration options. The RMS managed to satisfy this requirement from the outset, due to a combination of off-line and on-line computations.

5. **Transparency:** In D2.4, the stated requirement was that the criteria names and descriptions would be disclosed, but that the formulae computing their scores would not be. However, the RMS has been designed so that it would also be possible to disclose details regarding the scores of individual criteria.

   So, on the one hand the RMS is clearly designed to be part of the EU Community Platform (collecting information from its various components, as well as contributing its results to EurActory and PolicyLine), but on the other satisfies independent requirements exceeding the current requirements of the EU Community, as it may evolve as part of a follow-up project or a commercialization effort in way unforeseen in the User Requirements of the current project (D2.4).

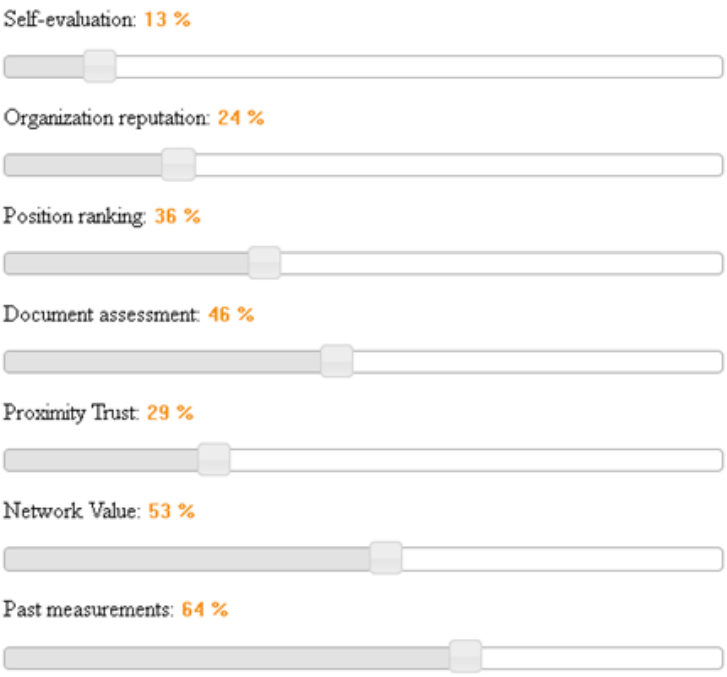

Figure 5: Weight Configuration Mockup Screen

3.4 Comparisons and Conclusions

The three reputation management systems examined earlier were purposefully chosen in order to frame the discussion of the EU Community RMS within a context of two very familiar albeit very different RMSs (eBay, Google Search) plus an RMS with which it has many similarities (as well as interesting differences) (Klout). The key similarities and differences of the four systems are summarised in Table 3. In more detail, this is how the EU Community RMS compares to the eBay, Klout and Google Search RMSs:

- Like the RMSs of Klout and eBay, EU Community's RMS also focuses on the reputation of individuals.
- Like Klout and Google Search, EU Community relies heavily on crawled content, but unlike them, not exclusively. It takes into consideration a number of criteria based on user input. It also includes an Offline Reputation criterion, which brings into the picture things happening outside the limited sphere of web content.
- Unlike the EU Community RMS, the Klout and eBay RMSs do not offer per topic reputation scores. The Google search RMS offers per query reputation scores, combining on-the-fly computed relevance scores and pre-computed page and site reputation scores.
- Like Google Search, and unlike Klout and eBay, EU Community supports subjective reputation criteria (or in Google's terminology, it supports personalisation). Google Search relies on things like past search, browsing history and geolocation information. The EU Community RMS relies on crawled social graph relations (see §5.5).

- The EU Community RMS is by design very customisable and it is up to the platform in which it is embedded to empower users and administrators to customise it using appropriate user interfaces. Google Search offer a degree of customisation but it is relatively limited in scope. Klout and eBay offer no customisability whatsoever, although, there is no reason to assume they are not customisable by design, but no user interface was ever provided to make turn their customisability potential into a feature available to users.

- Google and Klout's algorithms are considered trade secrets and especially in the case of Google secrecy is also considered a means of avoiding more black hat SEO attempts. Out of the three commercial RMSs examined, eBay's RMS stands out in that respect both because it is fully transparent but also because it is particularly simple. The EU Community RMS has been designed to be transparent, though it is up to the platform embedding it to expose as much information as seen fit for each application.

In summary, reputation calculations in eBay are transparent and use user-feedback, Google Search is personalisable and customisable and Klout lacks those features. At the same time, Klout is the closest of the three examined RMSs to it in term of its goals. Given that EU Community

1. targets a specific community of experts,
2. supports a mix of user-input and crawled reputation data,
3. computes reputation on a per-topic basis,
4. implements a variety of criteria not all of which linked to social media

it becomes obvious that the EU Community RMS has qualitative characteristics setting it apart from systems like Klout (where Jean-Claude Juncker may score lower than Justin Bieber or some blogger or other).

Table 3: RMS Comparison Matrix: eBay, EU Community, Google Search, Klout

FEATURE	**YES**	**NO**
REPUTATION FROM USER-SOURCED ASSESMENTS	eBay EU Community	Google Search Klout
REPUTATION FROM CRAWLED DATA	EU Community Google Search Klout	eBay
FOCUS ON PERSONS' REPUTATION	eBay EU Community Klout	Google Search (URLs)
SYNTHETIC REPUTATION CRITERIA	eBay (very simple) EU Community Google Search Klout	
COMPLEX COMPUTATIONS IN SYNTHETIC CRITERIA	EU Community Google Search Klout	eBay
SECRET ALGORITHMS FOR SYNTHETIC REPUTATION CRITERIA	Google Search Klout	eBay (simple, documented computations) EU Community (complex, open source algorithms)
PERSONALISATION / SUBJECTIVITY	Google Search EU Community	eBay Klout

4 Architecture, Performance, Communication and Integration

4.1 Off-Line and On-Line Processing

The Reputation Management System has to satisfy two important and opposing requirements:

1. Support criteria involving complex computations on large quantities of raw data.
2. Provide immediate real-time feedback to users configuring the reputation rating parameters.

In order to satisfy both these requirements the Reputation Management Component needs to perform:

- Off-line pre-computations of the ratings for individual criteria using pre-defined or administrator-defined logic.
- On-line computations using default or user-supplied options for the synthesis function, the result of which is computed on the basis of pre-computed reputation criteria ratings.

4.2 Integration with EurActory: Stored Procedures Collection vs. Communicating Component Design

4.2.1 The Stored Procedure Approach

The chosen approach is to:

(i) Encode the core RMS business logic as SQL stored procedures (which has the advantage of immediate read and update access to the relevant data).

(ii) Re-use the existing Drupal RESTful Web API for exposing the reputation scores (and separate calls for auxiliary intermediate scores).

This design preference is based on consideration of the following facts:

- The majority of primary reputation data come from data gathered by the Crawlers Component. The two exceptions are user-supplied data (e.g. self-assessments), gathered by EurActory, and sentiment analysis data, produced by the Opinion Mining and Sentiment Analysis Component. So, no single component whence some reputation data originates has all required data in order to calculate the reputation scores, but all three of them send their data over to CurActory, the common data repository.
- The single biggest overhead in an RMS that would communicate either directly with CurActory or via the project's Drupal RESTful API to obtain

the aforementioned data (excluding the unavoidable database server file I/O) would be the network communication overhead for getting data and storing back results; the requisite computations (per person, per document, per organisation, or any links between people and people, documents and documents, documents and people, and people and organisations) are trivial (they typically involve a very small number of primary arithmetic operations such as addition, multiplication and division). Therefore, for performance reasons alone, the obvious choice is to implement the core RMS business logic in the form of SQL stored procedures (CurActory is a database on a MySQL server). This is common practice for high-performance applications: keeping simple data-centric computation as close to the data as possible. Any solution involving taking data from CurActory, processing it elsewhere, and storing it back would be significantly less efficient.

- There are requirements for extreme customisation by administrators. Since SQL is by design an interpreted language (pre-processing and JIT optimisations aside), these become very easy to adhere to, if criteria scoring functions and synthesis functions are implemented in the common SQL server. In fact, it would be easy to even have the administrator define the so-called pre-defined functions mentioned here. Any auxiliary stored procedures and functions devised for use by the pre-defined criteria and synthesis functions would also be documented and available for use in the implementation of custom pre-defined criteria and synthesis functions.

- There is a requirement that reputation scores are refreshed daily.[4] This can easily be achieved using the MySQL scheduler e.g. by setting it up to run the main update loop every morning at 4am CET.

4.2.2 Another Possible Architectural Choice: The Separate Component Design

A different approach, which was eventually rejected, was to create a new external component, which would:

(i) take input from CurActory (possibly using the Drupal WebAPI rather than connecting directly to it)

(ii) process it

(iii) provide a web service for exposing its results ("The RMS must be inteconnectable, via web services, for providing reputation results per selected expert per topic [...].").

This approach has some serious disadvantages:

(i) It makes components that use the RMS output invoke two web methods, one on the RMS web service to get reputation data and one on a separate web method to the Drupal Web API to collect the

[4] While the requirement is reasonable, it is only feasible to satisfy it partially given the legal restrictions of the Linkedin API: criteria E and F which depend on Linkedin data may only be refreshed at the user's request at least for the part that pertains to Linkedin social graphs.

remaining information (or forces the duplication of CurActory data on the RMS database); currently all existing or planned components store data on and retrieve data from CurActory (directly or through the Drupal Web API) alone, so the RMS would be the anomaly defying established norm

(ii) If the RMS does not post data back to CurActory but instead stores it itself, we get into a scenario where there no longer is a single database, but two, one for the RMS and one for everything else; again this seems contrary to the overall design philosophy of the project

(iii) If the RMS does post data back to CurActory, the network overhead is not one but two data transfers that could have been avoided if the RMS core logic was placed in CurActory in the form of stored procedures and functions.

4.3 Implementation Technologies

As per the chosen architectural proposal, since CurActory is implemented as a MySQL database, MySQL Stored Procedures are being used to deliver the core RMS business logic. This provides for great flexibility and high performance due to direct access to the data the RMS needs to process both for the off-line and the on-line computations.

5 Reputation Criteria

The Reputation Criteria implemented in the third prototype of the Reputation Management System are listed below. For each reputation criterion, a (non-technical) description of the criterion is given, followed by classification information, a list of the sources from which raw data used in the computation originate, and information pertaining to the off-line and on-line pre-processing of CurActory data.

5.1 Criterion A - Self-evaluation

Description
It is a basic requirement of the EU Community platform that the user will be able to evaluate his/her own expertise in any number of topics. The user interface for this functionality is provided by EurActory. Typically, users will evaluate themselves on the topics most relevant to them, i.e. they will not provide a self-assessment for each and every topic. The absence of user-provided self-evaluation on a particular topic is to be interpreted as a zero-valued self-evaluation i.e. if the user does not claim to have any expertise in a topic, the system will understand this as a self-assessment of zero expertise on that topic. In the current version of EurActory, the self-evaluation user interface asks the user to select his/her areas of expertise. For areas chosen by the user, the self-assessment score recorded is 100. This user interface choice means that the only available self-assessment scores are 0 or 100 and not any value in between. The RMS can both support this or a more complex interface, say, with sliders allowing the user to specify exactly his level of expertise per area. However, the current user interface's simplicity helps from a usability perspective.[5] (It may be worth mentioning here that Linkedin took a similar approach with 'skills'.) Non-users for whom the Crawlers have provided data cannot have any per topic expertise self-assessments; in their case, there are two options: (i) that this criterion will not affect their reputation or (ii) that the value of the criterion will be 0. The latter option offers an advantage to users of the platform as opposed to non-users. Versions 1 and 2 of the Reputation Management Component takes this route.
Classification
▪ Per-person per-topic criterion ▪ Objective

[5] One of the external experts who reviewed the present deliverable recommended three choices: 0%, 50%, 100%. This could serve as the basis of a very usable user interface while avoiding the problems of our current all or nothing approach.

Sources
▪ EurActory
Off-line Precomputation
No pre-computation is required.
On-line Computation (for person *P,* topic *T*)
If person P has declared themselves an expert in topic *T* then return 100 else return 0.

5.2 Criterion B – Peer Assessment

Description
The peer assessment criterion allows experts to endorse each other in different areas of expertise. Unlike Linkedin's endorsement mechanism, for the purposes of this criterion, the reputation of the endorser is taken into account; therefore, an endorsement by two members with high reputation in a particular area will be more valuable by the endorsement of three members with low reputation. Two different design options for EurActory have been considered. The principle behind both was that EU Community must give users control of the areas of expertise they want to be associated with. The first was that EurActory would implement a user interface whereby one user may endorse another on any of the areas of expertise they have assessed themselves as being experts in. An alternative, more flexible design achieving the same goal was chosen for EurActory: endorsement is allowed on any area of expertise, but EurActory only displays in the user's profile those areas of expertise he/she has chosen. This design option has the additional advantage that peer recommendations can be used as input to a process of asking the user to consider adding additional areas of expertise to his/her profile. The RMS supports both options. It also supports rating the expertise levels (e.g. on a scale of 0 to 100), but there are no immediate plans to make use of this capability; in both aforementioned designs the value of an endorsement is fixed (100).
Classification
▪ Per-person per-topic criterion ▪ Objective
Sources

- EurActory

Off-line Precomputation

For each person *P* and topic *T:*

1. Let *S* be the sum of the adjusted values of the endorsements for *P* on *T*.
2. Find the RM_Person_Topic record for (*P,T).*
3. Set CriterionB_S=*S.*

For each topic *T:*

1. Let *M* be the maximum value of CriterionB_S.
2. For each person P
   a. Find the RM_Person_Topic record for (*P,T).*
   b. Set CriterionB=100**S/M.*

The adjusted value of an endorsement of person *P* by person *E* on topic *T* is v * r * f_1 * f_2 / e.

Where:

- *v* is the value of the endorsement (given the design choices mentioned above, *v*=100 for all endorsements)
- r is the endorser's i.e. *E*'s reputation on topic *T*
- f_1 = 1,if the endorser and the endorsee work for the same organization; otherwise f_1 = 1.5
- f_2 = 1,if the endorser and the endorsee work for the same organization; otherwise f_2 = 1.5
- e is the number of endorsements made by *E.*

On-line Computation (for person *P,* topic *T*)

1. Find the RM_Person_Topic record for the key (*P,T*)
2. Return Criterion_B

5.3 Criterion C – Business Card Reputation

Description

The Business Card assessment criterion assigns reputation to a person according to the information the EU Platform has regarding his/her job position(s).

Whereas Criteria A and B, depend solely on input given by users of the EU Community platform, Criterion C attempts to capture an aspect of real-life reputation: the reputation one's job position commands. It replaces two criteria described in D2.4 (User Requirements and Specifications) and implemented in the first version of the RMS (D3.2.1). As mentioned in D3.2.1, these two criteria, the Organisation Reputation and Position Reputation, were problematic, or more accurately, the fact that they were

initially designed to be separate was.

The problem arises when a person has multiple positions in different organisations. If the same person *P* has an important position at an unimportant organisation and an unimportant position in an important organisation, in version 1 of the RMS he/she would have received high reputation for both criteria. In the Conclusions and Future Work section this problem was identified and a solution involving a combined Position-Organisation Reputation criterion was promised. The Business Card criterion is exactly that promised solution.

There are two key ideas behind the Business Card criterion:

1. An organization (company, institution, NGO etc.) has a reputation that someone working for it can claim to the extent their position allows. This is analogous to how one typically evaluates another person's business card.
2. In cases of positions in different organisations, the one having the most reputation is chosen. This corresponds to a person giving their most flattering business card. This makes sure for example that a person with a reputable position does not lose reputation simply because he/she may also gain a less significant position in a less significant organization; the CEO of the BMW will not suddenly stop being the CEO of BMW if he/she decides to be the secretary of his local gardening club, nor will he distribute any business cards other than the BMW CEO ones for whatever other engagements he/she has at a meeting concerning the automotive industry.

These solve the aforementioned problem of version 1 of the RMS.

A third idea, is that unless a person has actively claimed expertise in a topic, no business card reputation will be given for him/her, or, more accurately, the criterion value will be 0.[6] Just like the case with business cards, if you do not want to use it to enhance your reputation, you do not.

Note: This idea is not integral to the design of the criterion, yet it helps focus our efforts on improving the results of the criterion for the platform's users in areas they have selected as their areas of expertise.

Classification

- Per-person per-topic criterion
- Objective

Sources

- Europa WhoIsWho
- LinkedIn
- EurActory (administrator-set scores)

[6] While there could be a separate option for turning this criterion on or off, in our design self-assessment is used for this purpose.

Off-line Precomputation

For each person *P* and topic *T:*

1. Let *S* be the maximum organisation-position score for person *P* on topic *T*
2. Find the RM_Person_Topic record for (*P,T)*
3. Set CriterionC_Raw=*S*

For each person *P* and topic *T:*

1. Let *M* be the maximum organisation-position score on topic *T*
2. Find the RM_Person_Topic record for (*P,T)*
3. If *P* is an expert in *T* according to self-assessment, then set CriterionC= 100*CriterionC_Raw*/M,* otherwise set CriterionC=0.

The organisation-position score of person *P* on topic *T* for a position with job title *J* held in organization *O* is defined thus:

$$o(O,T)*t(J,T)/100$$

Where

- o(*O*,*T*) =the reputation value for organisation *O* on topic *T* retrieved from the RM_Person_Topic record for (*P,T)*
- t(*J,T*) = the value of the automatically derived reputation score for that person's job title associated with the current position, unless an administrator-override reputation value exists in which case it is returned instead. The automated value is computed as follows: s(J) * $1.1^{\min(3,r(J,T))}$.
- s(J), the seniority of *J,* is a score corresponding to the seniority of the job title assigned by seeing if *J* matches any known patterns and assigning the corresponding reputation score; 0 is returned if no match is found. For example, a "Head of Sector" will receive a score of 50, a "Head of Unit" will receive a score of 60, a "Director" a score of 70, a "Director-General" a score of 80, a Commissioner a score of 90 and the President of the Commission a score of 100.
- r(J,T), the relevance of *J* to *T*, is a textual similarity metric calculated as the number of key-words shared between *J* and *T*.[7]

[7] For example, if the topic *T* is 'Energy Efficiency' and the job title is 'Director-General, DG ENER – Directorate-General for Energy' *r(J,T)=1* whereas if the job title is `Director, Dir C — Renewables, Research and Innovation, Energy Efficiency, DG ENER – Directorate-General for Energy' *r(J,T)=3;* in practice this means that when both seniority and relevance are considered the first position will have a score of 80*1.1=88 while the second will have an even higher score, 70*1.1*1.1*1.1=93.17.

On-line Computation (for person *P,* topic *T*)
1. Find the RM_Person record for the key (*P,T*) 2. Return Criterion_C

5.4 Criterion D – Document assessment

Description
This criterion assigns a reputation rating on the basis of the document reputation indicators that documents produced by the person have received. The Document Assessment criterion is not a criterion about who someone is and what position they hold (Criterion C), nor about who they know (Criteria E and F), nor about what they (Criterion A) and others (Criteria B and G) think they are an expert in; it is a criterion that evaluates a person's contribution to the past and ongoing EU policy processes. As such, it does not discriminate against new users not having the benefit of an important position and/or connections and offers them a chance to organically increase their reputation as a result of enriching the platform for the benefit of the community. The criterion's specifications and implementation, starting with version 2 of the RMS, follow the requirements set out in D2.4, relying on the platform's document assessment scores. This was not possible in version 1 of the RMS as PolicyLine was not available and the ground rules for how the two components would collaborate had not been set. For each document, PolicyLine enables users to answer a short questionnaire assessing issues of quality, relevance and position endorsement. The weight of their answers is determined by their reputation score. Once a questionnaire is submitted, PolicyLine will update both a document score it will use for its own purposes, but will also store the three questionnaire scores, which the RMS can use for the purpose of this criterion. **Note**: The RMS does not use the document score as computed by PolicyLine because that has a different purpose and takes into account also the author's reputation whereas the RMS only wants an evaluation of the document itself. Moreover, it requires the three scores separately (quality, relevance, endorsement) as this offers more flexibility in how they are used. More specifically, the RMS makes a distinction between the two scores that are relevant to the document, namely the quality and endorsement scores, and the relevance score. Whereas quality and endorsement as attributes of the document itself affect the reputation of the document author(s), the relevance score needs to be treated differently. Relevance assesses, as its name suggests, how relevant the document is to a given context, namely a given policy process. It is not an attribute of the document itself but of the link between the document and the policy process. It is possible that the same document, if it is for example a scientific paper or a report providing

relevant facts, to be linked to more than one policy process. Moreover, it is possible that someone other than the author(s) of the document links the document to one or more policy processes.

So:

1. There may be more than one authors for a document and its quality and endorsement scores should affect their reputation,
2. the document may be linked to more than one policy processes, possibly by different users, but only one use is responsible for linking it a document to a policy process and the relevance score should affect his/her reputation (not the authors').

Two more questions to be answered:

1. What is the relevant weight of quality, relevance and endorsement for the purposes of this criterion? It was decided to give them equal weights; this is not according to D2.4 (where quality and relevance are given equal weights but endorsement is valued as six times more important), but D2.4 is focused on document assessment for the purposes of PolicyLine.
2. How is the case of multiple author's handled? Whereas it would be possible to split the reputation arising from a document equally between its authors, we have for the time being decided to assign it in full to all of them.

The document score, like all the other criteria scores is normalised, i.e. in the range of 0 to 100; this is achieved by summing up all the relevant document scores for each user and each document and then dividing that score, the raw score, by the maximum score and multiplying by 100 to get the normalised score.

Note: This normalization method can be applied to any criterion with a variable range of raw scores.

Classification

- Per-person per-topic criterion
- Objective

Sources

- PolicyLine

Off-line Precomputation

For each person *P,* each topic *T*:

1. Let *Dq* be the sum of the quality scores of documents (co-)authored by *P* linked to policy processes under topic T.
2. Let *D*e be the sum of the endorsement scores of documents (co-)authored by *P* linked to policy processes under topic T.
3. Let *Dr* be the sum of the relevance scores of documents linked by *P* to policy processes under topic T.

4. Update the RM_Person_Topic record for the key (*P,T)*, setting the field Criterion_DRawScore= *Dq*+ *Dr*+ *De.*

For each topic *T*:

1. Let *M* be the maximum value of Criterion_DRawScore in RM_Person_Topic.
2. Update all entries for topic *T* in RM_Person_Topic as follows: Criterion_D = 100×Criterion_DRawScore/*M.*

On-line Computation (for person *P,* topic *T*)

1. Find the RM_Person_Topic record for the key (*P,T*).
2. Return Criterion_D.

5.5 Criterion E – Proximity Trust

Description

The Proximity Trust Criterion is not an objective reputation criterion, but rather a subjective relevance criterion that allows a user's assumed personal bias towards being interested more in results involving his/her own acquaintances to be taken into consideration by the system. It is (reasonably) assumed that the user is more interested in the documents and profile changes of his/her acquaintances, the closer these acquaintances are to him/her, as can be determined by social graph distances (LinkedIn, Twitter) or organizational charts (WhoIsWho/LinkedIn).

Different scores can be given to connections according to the proximity and the type of connection. The RMS employs a proximity trust point system, computed on-line, awarding respectively 35 points for a 1st degree connection on LinkedIn, 30 for a 'follows' relation and 5 for a `is-followed-by' relation on LinkedIn and 30 for being colleagues (under the broad definition of working for the same company or institution according to WhoIsWho or LinkedIn).

Note: Currently, the RMS does not use Linkedin as a data source as the LinkedIn Public API recently restricted access to the data the RMS was using.

Classification

- Per-person criterion
- Subjective

Sources

- CurActory (including job position data from Europa WhoIsWho and Linkedin)
- LinkedIn (not currently due to recent API restrictions)

▪ Twitter
Off-line Precomputation
For each person *P* and person *U*, 1. Let s = 100 if P=U, or otherwise: (If *P* and *U* are 1st degree connections on LinkedIn THEN 35 ELSE 0) + (If *P* follows *U* on Twitter THEN 5 ELSE 0) + (If *U* and *P* are colleagues according to CurActory THEN 30 ELSE 0) 2. Update the RM_Person_Person record for the key (*P,U)*, setting the field Criterion_E= *s.*
On-line Computation (for persons *P, U*)
Note: P is the person about whom the score is being computed, whereas *U* is the person for whom the score is being reported i.e. *U* is the current user. Technical detail: *U* is person id, not a user id. 1. Find the RM_Person_Person record for the key (*P,U)* 2. Return Criterion_E.

5.6 Criterion F – Network Value

Description

This is an objective reputation criterion that allows the system to take into consideration the influence of one's connections on his/her reputation. For instance, if a certain energy expert is followed on Twitter by over half the current MEPs, this is very indicative of his/her reputation and influence. Someone's raw network value score is computed as the sum of the adjusted reputation values of that person's network connections. An adjusted reputation value is a reputation value multiplied by the adjustment factor corresponding to the relation between the two individuals.

Relation Type	Adjustment Factor
Twitter 'is-followed-by' relation	100%
LinkedIn '1st degree connection' relation	50%

Note 1: The Twitter 'follows' relation is not taken into consideration in order to avoid reputation-metric manipulation by means of following of influential figures. Besides, as anyone can follow anyone, the `follows' relation does not say much about the reputation of the follower.

Note 2: The Twitter 'is-followed-by' relation is considered more important than the LinkedIn 1st degree connection both because it is asymmetric and because it depicts an interest in what the person being followed has to say, as opposed to a real-life social connection.

Note 3: In version 1 of the RMS no use of Linkedin data was made, as we had stumbled upon a legal constraint in the LinkedIn API Terms of Use. Further investigation revealed that this constraint can be accommodated, by explicitly gaining the user's consent for using Linkedin connections data to update his/her E and F reputation scores on the spot and by deleting such data after they had been used for this purpose. It was decided that until the appropriate user-interface asking for such permission was put into place, the RMS would not have make use of this kind of data from Linkedin, though the infrastructure to handle them is already in place. Further developments, namely the change in Linkedin's policy for its Public API, meant that the infrastructure remains unused. As it may be possible to reach a commercial agreement in the future, we will not disregard the already built-in support for Linkedin data in our discussions below.

Note 4: The LinkedIn 2nd, 3rd (and 4+th) degree connections (as well as common group membership connections) are taken to be possibly incidental, so only 1st degree connections are to be taken into consideration, if possible, as only they represent a conscious statement of both parties about their social connection.

Note 5: The adjustment factors could be administrator and user configurable, but in the current version they are not.

Note 6: As in the case of a two-way connection between persons A and B, the network value of A can influence the network value of B and vice versa. If the reputation metrics used in the computation are the vanilla reputation metrics incorporating this criterion, each reputation value update will result in higher and higher raw network values for both A and B; in order to avoid that, a reputation metric excluding Criterion F must be used. This problem was mentioned in D3.2.1 and has been addressed in the later two versions of the reputation management system; the reputation score taken into consideration is the same as the system reputation score except that the weight of criterion F is assumed to be 0 irrespective of the system default weight.

Note 7: In many Reputation Management Systems, including Klout and PageRank (a component of the synthetic Google Search RMS but possible to also examine as an autonomous RMS), there is a concept of reputation dilution. In the context of this criterion, if person *A* had a reputation *R* and followed person *B,* whereas currently this would benefit *B*'s raw score by 100%×*H,* if reputation dilution was implemented the question of how many others *A* follows would become a factor. In the simplest implementation of reputation dilution, if person *A* had a reputation *R* and followed *F* accounts, including person *B's* account*,* this would benefit *B*'s raw score by 100%×*H/F.* A reputation dilution option is being considered for the third version of the EU Community RMS.

As the raw network value does not necessarily fall in the range of 0 to 100, the same normalization process described for Criterion D is used for obtaining the normalised network value from the raw network value; the difference is that the Twitter and Linkedin scores are computed separately

allowing the relative weighting decision to take place during the on-line part of the computation. Whereas the weights are currently fixed, this could change in a future version of the RMS. **Note**: Currently, the RMS does not use Linkedin as a data source as the LinkedIn Public API recently restricted access to the data the RMS was using.
Classification
▪ Per-person, per-topic criterion ▪ Objective
Sources
▪ Twitter ▪ LinkedIn (not currently due to recent API restrictions)
Off-line Precomputation
For each person *P* and each topic *T*: 1. Let *F1* be the sum of the system reputations of *P*'s Twitter followers 2. Let *F2* be the sum of the system reputations of *P*'s 1st degree LinkedIn connections 3. Update the RM_Person_Topic record for the key (*P*,*T*) setting the fields Criterion_F1=*F1,* Criterion_F2=*F2* For each topic *T*: 1. Let *MF1* be the maximum value of Criterion_F1 in RM_Person_Topic 2. Let *MF2* be the maximum value of Criterion_F2 in RM_Person_Topic 3. Update all entries for topic *T* in RM_Person_Topic as follows: Criterion_F1 = 100×Criterion_F1/*MF1*, Criterion_F2 = 100×Criterion_F2/*MF2*
On-line Computation (for person *P* and topic *T*)
1. Find the RM_Person_Topic record for the key (*P*,*T*) 2. Return (τ×Criterion_F1+ λ ×Criterion_F2)/(τ+λ) where τ=100%, λ=*50%*

5.7 Criterion G – Offline Reputation

Description
Given that not everybody with an active role in European policy processes necessarily has a strong, or indeed any, presence in social media, the EU Community RMS has an offline reputation criterion that attempts to

capture this aspect of a person's reputation. This is a criterion that involves manual labour and a special interface for the designated small number of support-staff that will be responsible for the relevant task. This criterion allows for many factors to be taken into consideration overcoming technical complications having to do with the availability and processability of on-line information. In D2.4 the Offline Reputation criterion was envisaged as a criterion where reputation would be gathered both by participants of events and by asking the question to the participant: 'Who else should have been here?'. These are concrete verifiable data points that can be dealt with algorithmically to produce offline reputation scores (or be components of offline reputation scores). More details about the way Offline reputation has been computed in the current scores are given in Section 10.4.
Classification
▪ Per-person, per-topic criterion ▪ Objective
Sources
▪ EurActory (special interface for EU Community's support staff)
Off-line Precomputation
The criterion is not computed by the RMS, but rather by methods outside it and the scores are then made available to the platform via a special support-staff interface provided by EurActory.
On-line Computation (for person *P* and topic *T*)
1. Find the RM_Person_Topic record for the key (*P*,*T*) 2. Return Criterion-G

5.8 Criterion H – Past Measurements

Description
Other criteria may have values that change from one day to another. If that change is drastic there will be abrupt fluctuations in a person's reputation. Criterion H allows historical reputation values to smoothen the fluctuation and reflect the intuition that a person's reputation today is not the result of the current state of affairs only (e.g. current position) but also

of his/her reputation in the past.

Note: Exactly which time frame interests the consortium is a matter to be discussed, but it seems that a reasonable proposal would be to take into consideration the reputation of January 1st of four years ago, of twelve months ago, of the 1st of six months ago, and of the 1st of the previous month, attach different (user/admin configurable) weights to each and take the weighted average as the value for this criterion.

Classification

- Per-person, per-topic criterion
- Objective

Sources

- CurActory (historical Reputation Management System data)

Off-line Precomputation

The system reputation scores are kept for the 1st day of each month for each person and topic.

For each person *P* and each topic *T:*

1. Let *H1* be the archived system reputation of *P* on *T* on January 1st of four years ago
2. Let *H2* be the archived system reputation of *P* on *T* on the 1st of twelve months ago,
3. Let *H3* be the archived system reputation of *P* on *T* on the 1st of six months ago,
4. Let *H4* be the archived system reputation of *P* on *T* on the 1st of last month
5. Update the RM_Person_Topic record for the key (*P*,*T*) setting the fields Criterion_H1=*H1,* Criterion_H2=*H2,* Criterion_H3=*H3,* Criterion_H4=*H4.*

On-line Computation (for person *P* and topic *T*)

1. Find the RM_Person_Topic record for the key (*P*,*T*)
2. Return
$$\frac{\text{Criterion\_H1} * \text{CriterionH1Weight} + \cdots + \text{CriterionH4} * \text{CriterionH4Weight}}{\text{CriterionH1Weight} + \cdots + \text{CriterionH4Weight}}$$

Note: The values of the weights could be administrator-set or even user-set, as they are only used in the on-line computation. They are currently fixed:

CriterionH1Weight=25,

CriterionH2Weight=25,

CriterionH3Weight=25,

CriterionH4Weight=25.

5.9 Criterion I – Outcome and Duration Predictions Score

Description

PolicyLine is a tool to be used by EU Policy Experts. Experts are invited to inform the platform of their estimates about the outcome and duration of ongoing legislative procedures. Experts will not always be right and some experts may be right more often than others; this is going to be reflected in their prediction scores. These scores will be computed by the Policy Component (upcoming deliverables D4.3.2 and D4.4.2) and imported by the Reputation Management System.

Note: Whereas there are ways to make the relevant scores topic-specific, the decision was to have them topic-agnostic; this will allow experts, including generalists, to compete across the board rather than only specific areas of EU politics.

Classification

- Per-person criterion
- Objective

Sources

- Policy Component

Off-line Precomputation

For each person *P:*

1. Let *I_1* be P's score for outcome predictions, M_1 the maximum score for outcome predictions and m_1 the minimum score for outcome predictions
2. Let *I_2* be P's score for outcome predictions, M_2 the maximum score for outcome predictions and m_2 the minimum score for outcome predictions
3. Update the RM_Person record for the key *P* setting the fields Criterion_I1=*I1*100/(M_1-m_1+1),*
   Criterion_I2= *I2*100/(M_2-m_2+1).*

On-line Computation (for person *P* and topic *T*)

1. Find the RM_Person record for the key *P*
2. Return

$$\frac{\text{Criterion\_I1} * \text{CriterionI1Weight} + \text{CriterionI2} * \text{CriterionI2Weight}}{\text{CriterionI1Weight} + \text{CriterionI2Weight}}$$

Note: The values of the weights could be administrator-set or even user-

set, as they are only used in the on-line computation. They are currently fixed and their values are:

CriterionI1Weight=50,

CriterionI2Weight=50.

6 Synthesis Functions

A synthesis function is a function that takes the individual per-criterion metrics and produces a single reputation metric. It is possible for administrators to add synthesis functions of their choosing. The pre-defined synthesis functions are listed below.

6.1 Default

Description
The weighted average of the criteria scores.
User Configuration
▪ Weights per criterion **Note:** It is envisaged that the user will have available a user interface for de-activating the weighting of criteria. This will not be supported by the RMS in terms of an additional flag parameter to the synthesis function as the same effect can be obtained by using 1 as the weight of all criteria.
Implementation
$\frac{\text{CriterionA}*\text{CriterionAWeight}+\cdots+\text{CriterionG}*\text{CriterionGWeight}}{\text{CriterionAWeight}+\cdots+\text{CriterionGWeight}}$

6.2 Best *N* (e.g. Best 3)

Description
The weighted average of the best N weighted criteria scores. Example: Suppose the user chooses N to be 3 and disables the weighting of criteria (see User Configuration below). For person X, the 3 best criteria may be (D:40) (A:35) and (G:15) (average: 30) whereas for person Y the 3 best criteria may be (C:10) (B:20) and (D:30) (average: 50). So person Y will have a higher reputation than person X using this synthesis function even if on average person X fairs better in terms of individual per-criterion scores.
User Configuration
▪ *N* ▪ Weights per criterion **Note:** It is envisaged that the user will have available a user interface for de-activating the weighting of criteria. This will not be supported by the

RMS in terms of an additional flag parameter to the synthesis function as the same effect can be obtained by using 1 as the weight of all criteria.
Implementation
1. Map each criterion *i* to a row containing the score for Criterion *i*, the weight for Criterion *i* and the product of the score for Criterion *i* × the weight for Criterion *i* 2. Put those rows in ascending order of the third column i.e. he score for Criterion *i* × the weight for Criterion *i* 3. Calculate S the sum of the third column of *N* rows starting with row 1 4. Calculate W the sum of the second column of *N* rows starting with row 1 5. Return S / W

6.3 Worst N (e.g. Worst 3)

Description
The weighted average of the worst N weighted criteria scores.
User Configuration
▪ *N* ▪ Weights per criterion **Note:** It is envisaged that the user will have available a user interface for de-activating the weighting of criteria. This will not be supported by the RMS in terms of an additional flag parameter to the synthesis function as the same effect can be obtained by using 1 as the weight of all criteria.
Implementation
1. Map each criterion *i* to a row containing the score for Criterion *i*, the weight for Criterion *i* and the product of the score for Criterion *i* × the weight for Criterion *i* 2. Put those rows in descending order of the third column i.e. he score for Criterion *i* × the weight for Criterion *i* 3. Calculate S the sum of the third column of *N* rows starting with row 1 4. Calculate W the sum of the second column of *N* rows starting with row 1 5. Return S / W

6.4 Middle N (e.g. Middle 3)

Description
The average of the middle N weighted criteria scores.
User Configuration
▪ *N* ▪ Weights per criterion **Note:** It is envisaged that the user will have available a user interface for de-activating the weighting of criteria. This will not be supported by the RMS in terms of an additional flag parameter to the synthesis function as the same effect can be obtained by using 1 as the weight of all criteria.
Implementation
1. Map each criterion *i* to a row containing the score for Criterion *i*, the weight for Criterion *i* and the product of the score for Criterion *i* × the weight for Criterion *i* 2. Put those rows in descending order of the third column i.e. he score for Criterion *i* × the weight for Criterion *i* 3. Let F=floor((8-N)/2) + 1 4. Calculate S the sum of the third column of *N* rows starting with row *F* 5. Calculate W the sum of the second column of *N* rows starting with row *F* 6. Return S / W

7 Objective and Generic Reputation

Whereas the RMS is designed to compute a reputation score that differs from topic to topic and from observer to observer (whereby observers are the EU Community Platform users), namely a subjective per person per topic score, on occasion that specificity is not merely unnecessary but unwanted. What follows, is a specification pertaining to how the cases of having no current user (specified by passing 0 as the current user person id) or no topic (specified by passing 0 as the topic id) as inputs are dealt with by the RMS.

7.1 Objective Reputation

One of the most distinguishing features of the RMS is that it supports subjective criteria i.e. criteria whereby for a user an expert receives a certain score and for a different user a different one. Yet there are possible usage scenarios where subjectivity is undesirable:

- The per topic ranking is objective, as there needs to be a single EurActory ranking of experts per topic, not multiple rankings whereby experts have certain ranks when seen by one user and a different one when seen by another user; this unique (per topic) ranking needs to be based on an objective score i.e. on a score that does not subscribe to the view that reputation is in the eye of the beholder.
- The computation of the system reputation value for historical record keeping, as required by Criterion H, is topic sensitive, but not user-sensitive i.e. an objective score is required.
- The computation of the weights of topic for computing Generic Reputation (see Section 7.2) is also based on objective scores.

In order for the RMS to eliminate observer-based variability and produce an objective score, its subjective criteria (currently, Criterion E) need to be excluded from the reputation score computation.

This is achieved by passing (instead of a valid user identifier) 0 as the value of the current user parameter. In this case, the weight of Criterion E is also set to 0 (irrespective of what weight may have been passed as a parameter to the RMS scoring function) thus ensuring Criterion E is not considered, letting the score be determined entirely on the basis of the remaining criteria.

A similar approach could also be used when no topic is specified (ignore all topic-sensitive reputation criteria). Nevertheless, this would leave out important information from consideration of a person's reputation. So a different solution is used there, as described below.

7.2 Generic Reputation

The need to support the calculation of a score without reference to a topic arises in cases where the user is interacting with EurActory or PolicyLine in a context where no topic has been explicitly or implicitly been selected.

One way around that, would have been to ignore all topic-sensitive criteria (treat them as having zero weight) and produce a result based only on the remaining ones. This was already less than ideal in the first version of the RMS; it would have been even less ideal in the second or the current, third, version in which the majority of the criteria are topic-sensitive.

Previously, the RMS supported a special dummy topic id, 0, which would get, for each expert, the maximum reputation he/she has in the other (i.e. the real) topics, for each reputation field in the RM_Person_Topic table. This was implemented as a final step in the pre-computation of reputation scores. As a result, the on-line computation did not treat the case of no topic (topic=0) in any special manner.

In the third version of the RMS, a more sophisticated method for computing a generic reputation score is in place. Whereas the earlier method considered only the expert's best score across all topics as his/her generic reputation, thus clearly favouring specialisation, the third version of the RMS provides a more balanced approach considering all topics in the platform.

Step 1: For each topic *t*, compute its weight W_t as the sum of all expert scores in that topic.

Step 2: Compute S, the sum of all the topic weights.

Step 3: For each expert, set his/her generic reputation score to be the weighted sum of his/her topic-specific reputation scores divided by S.

Assuming all weights are equal, the new way of computing generic reputation would mean that an expert having good scores on all topics would score higher than an expert with an excellent score on a single topic and near zero scores on the remaining topics. This is in contrast with the earlier versions' generic reputation computation that would favour the latter expert.

But weights need not be equal. Indeed, this is one of the more interesting aspects of the new algorithm.

If there were only two topics, A and B, and the overall average reputation in A is much higher than the overall average reputation in topic B, an expert with a high score on A and a low score on B, would have a higher generic reputation than another expert having the exact same scores reversed (i.e. scoring better on the lower weighted topic). This too is in contrast with the earlier versions' generic reputation computation that would assign the same generic reputation score to the two experts.

What is interesting is that the sum of reputations in a topic (or, equivalently, the average reputation in the topic) is used as a weight for scores in that topic, thus determining its significance in the generic reputation score calculation. Practically, this means that topics where EU Community users are more active will be considered more important in the computation of the general reputation score. The intuition here is that the more important a topic is within the platform, the more weight it will be given when generic scores are computed.

8 Integration

The Reputation Management System is not a user-facing component; rather, its purpose is to provide reputation information to such components of the EU Community Platform, namely EurActory and PolicyLine in a convenient and efficient manner.

Both EurActory and PolicyLine have direct access to CurActory, the common EU Community Platform database. By the RMS's design, they both remain entirely agnostic to the RMS's design, data and operations except for:

- the fact that they store information that is used by the various RMS criteria (see above)
- and the fact that they have access to two functions made available by the RMS: RM_SCORE_WITH_FULLPARAMETERS and RM_SYSTEM_SCORE (see below).

RM_SCORE_WITH_FULLPARAMETERS		
Input Parameter 1 Name _PersonId	**Input Parameter 1 Type** LONG	**Input Parameter 1 Documentation** This is the id of the person for which the reputation score for the topic specified by the second argument will be computed.
Input Parameter 2 Name _TopicId	**Input Parameter 2 Type** LONG	**Input Parameter 2 Documentation** This is the id of the topic for which the reputation score for the person specified by the first argument will be computed. Together, parameters 1 and 2 specify the two key inputs for the reputation scoring question in the EU Community RMS: (i) Whom are we evaluating the reputation of? (ii) On what topic are we evaluating his/her reputation?

		When a per-person rather than a per-person per-topic score is required, the value 0 must be passed to the _TopicId parameter, as discussed in §**Error! Reference source not found.**.
Input Parameter 3 Name _CurrentUserPersonId	**Input Parameter 3 Type** LONG	**Input Parameter 3 Documentation** This is the id of the user on behalf of whom the reputation score is being requested. This parameter is required because the EU Community Reputation Management System is capable of taking into consideration subjective (i.e. user-sensitive) factors (such as social network proximity – see §5.5) in computing person-topic reputation scores. When requesting an objective (i.e. non user-sensitive) person-topic reputation score, the value 0 must be passed to the _CurrentUserPersonId parameter, as discussed in § 7.2.
Input Parameter 4 Name _Synthesis	**Input Parameter 4 Type** VARCHAR(30)	**Input Parameter 4 Documentation** This parameter determines the synthesis function to be used for computing the synthetic reputation score. It may take one of the following values: ▪ '$DEFAULT' ▪ '$BEST-N' ▪ '$MIDDLE-N' ▪ '$WORST-N' These values correspond to the synthesis functions specified in Section 0. For more details, see the documentation of the _N_Criteria parameter.

Input Parameter 5 Name	**Input Parameter 5 Type**	**Input Parameter 5 Documentation**
_N_Criteria	INTEGER	This parameter specified the number N i.e. the number of criteria that will be taken into consideration by the synthesis function specified via the _Synthesis parameter when the latter is not '$DEFAULT', but rather one of the following: ▪ '$BEST-N' ▪ '$MIDDLE-N' ▪ '$WORST-N' Together parameters _Synthesis and _N_Criteria specify the synthesis function to be used in determining the synthetic reputation of _Person in _Topic as perceived by _CurrentUserPersonId.
Input Parameters 7-14 Names CriterionAWeight CriterionBWeight CriterionCWeight CriterionDWeight CriterionEWeight CriterionFWeight CriterionGWeight CriterionHWeight CriterionIWeight	**Input Parameters 7-13 Type** DOUBLE	**Input Parameter 7-13 Documentation** The weights to be given, respectively, to the reputation scores of Criteria A to H in determining the synthetic reputation of _Person in _Topic as perceived by _CurrentUserPersonId in accordance with the synthesis function specified by _Synthesis and _N_Criteria. Example: If for the given _Person the reputation score according to Criterion A in _Topic as perceived by _CurrentUserPersonId is 80 and _CriterionAWeight is 0.1, then the Criterion A score that will be considered by the synthesis function will be 80×0.1=8. Note: There are no algorithmic requirement for weights to fall in any particular range, nor to add up to a particular number, but assuming they will be in

		the range 0 to 1 and add up to 1 makes this discussion easier. Every individual criterion score is guaranteed to be in the range 0 to 100, and the synthetic score will also be in the same range.
	Output Type DOUBLE	**Output Documentation** The synthetic reputation score of _Person in _Topic as perceived by _CurrentUserPersonId in accordance with the synthesis function specified by _Synthesis and _N_Criteria, taking into consideration the _CriterionAWeight … _CriterionIWeight weight parameters.

RM_SYSTEM_SCORE		
Input Parameter 1 Name _PersonId	**Input Parameter 1 Type** LONG	**Input Parameter 1 Documentation** This is the id of the person for which the reputation score for the topic specified by the second argument will be computed.
Input Parameter 2 Name _TopicId	**Input Parameter 2 Type** LONG	**Input Parameter 2 Documentation** This is the id of the topic for which the reputation score for the person specified by the first argument will be computed. Together, parameters 1 and 2 specify the two key inputs for the reputation scoring question in the EU Community RMS: (iii) Whom are we evaluating the reputation of?

		(iv) On what topic are we evaluating his/her reputation? When a per-person rather than a per-person per-topic score is required, the value 0 must be passed to the _TopicId parameter, as discussed in §Error! Reference source not found..
Input Parameter 3 Name _CurrentUserPersonId	**Input Parameter 3 Type** LONG	**Input Parameter 3 Documentation** This is the id of the user on behalf of whom the reputation score is being requested. This parameter is required because the EU Community Reputation Management System is capable of taking into consideration subjective (i.e. user-sensitive) factors (such as social network proximity – see §5.5) in computing person-topic reputation scores. When requesting an objective (i.e. non user-sensitive) person-topic reputation score, the value 0 must be passed to the _CurrentUserPersonId parameter, as discussed in § 7.2.
	Output Type DOUBLE	**Output Documentation** The synthetic reputation score of _Person in _Topic as perceived by _User in accordance with the *system default* synthesis function and weight parameters. Given that no administrative interface is available for setting these system defaults they are currently fixed, in accordance with the requirements in D2.4 (with the addition of the

		Network Value and Past Measurements criteria for which weights were chosen irrespective of D2.4 as they were not included there) i.e. the value returned is:
		`RM_SCORE_WITH_FULL_PARAMETERS(`
		`_PersonId,`
		`_TopicId,`
		`_CurrentUserPersonId,`
		`'$DEFAULT', /*SYNTHESIS*/`
		`9, /*N*/`
		`5, /*A.Self-Assesment*/`
		`10, /*B.Self-Assesment*/`
		`10, /*C.Business-Card*/`
		`15, /*D.Documents*/`
		`15, /*E.Proximity*/`
		`15, /*F.Network*/`
		`30, /*G.Offline*/`
		`15, /*H.Past*/`
		`15 /*I.Predictions*/)`
		Note: Even if the sum of the weights exceeds or does not reach 100, this does not affect the range of the `RM_SYSTEM_SCORE` function which will be 0 to 100 as `RM_SCORE_WITH_FULL_PARAMETERS` performs the necessary computations to guarantee the score will fall on that range, as already explained above.

9 Weights Adjustment Methodology

The weights given to the various criteria determine the emphasis given to specific aspects of reputation and affect the scores the RMS generates. Determining what the weights will be can be crucial factor in the success or failure of the RMS to provide a credible estimation of policy experts' reputations.

9.1 Initial Methodology

The User Requirements deliverable (D2.3) provided a set of weights for the criteria foreseen at that time, based on feedback from potential users of the EU Community Platform. Given that weights for the RMS criteria were initially derived from D2.3, the initial weight selection methodology involved translating the importance of criteria as perceived by a number of potential users and/or external experts into weights. It would have been possible to base weights on WP3 External Experts opinions regarding their perceived importance of different criteria.

From the outside, this seems to be a defensible choice of methodology. It is practical, in the sense that it is easy to apply and it would seem to guarantee the credibility of the synthetic score, to the extent the criteria work as they should and the respondents are credible experts/potential users.

9.2 Revised Methodology

Whereas it is an interesting experiment to seek human expert knowledge and intuitions when designing the RMS and when deciding the weights of its criteria, at the end of the day, it is the end result of the RMS that is evaluated and is judged to be convincing or not. In prior versions of the RMS, the emphasis was on implementation. The initial weight selection used in both the first two versions was deemed sufficient, simply on the grounds of being a reasonably, defensible methodology.

The first two prototypes were experiments, but the third version was meant to be exploited as the basis of a commercial service, one that can be advertised and sold to a demanding and highly opinionated audience. In the third version, emphasis was on end results. In order to provide a credible ranking, policy domain expertise is now used to evaluate exactly how credible the actual ranking is.

There was a move from a process that looked at criteria in the abstract to a process that:

- looked at the results of the weighting applied,
- had a transfer of feedback from policy domain experts to the RMS technical team who adjusted the criteria weights so that feedback could be obtained again and
- was repeated until a point where the adjustment of the weights was deemed not to require further improvements.

Table 4: RMS Criteria Weights

	D2.3 User Requirements	**D3.2.2** Second Prototype	**D3.2.3** Third Prototype
Self-Evaluation	10	10	5
Peer Assessment	15	15	10
Business Card (Position&Title)	10	10	10
Document Assessment	25	25	15
Proximity Trust (PolicyLine Only)	15	15	15
Network Value	-	15	15
Offline Reputation	25	25	30
Past Measurements		10	15
Predictions Score	-	-	- /15[8]
Total (EurActory)	**85**	**110**	**100**
Total (PolicyLine)	**100**	**125**	**110**

Note 1: As there is only one EurActory rankings per topic, i.e. the ranking does not depend on who the user currently looking at it is, the Total of the weights used in EurActory does not include the Proximity Trust criterion weight. As a user will be more interested in documents authored by individuals they know, the Proximity Trust criterion is used in PolicyLine, hence the different weights scores totals in EurActory and PolicyLine.

Note 2: The total reputation score in both EurActory and PolicyLine is always in the range of 0 to 100, irrespective of the total of the weights (Chapter 8).

The approach outlined above was applied ahead of the high-profile project dissemination event on 24 February 2016 at the European Parliament where the Energy Union rankings were revealed. The criteria weighting has not changed since then, but a new criterion has been added ("Prediction Scores") and when we decide to introduce it into the RMS ranking and calculations an appropriate

[8] There have been two major releases of the third version of the RMS. The first release served as the basis for the first validated rankings unveiled in the highly successful events. The Predictions Score criterion was added (with a weight of 15) in the second release of the RMS, a forward looking release that aims to serve as the basis of further commercial and research exploitation.

weight for it will have to be assigned, possibly as part of an overall recalibration of the RMS.

It is worthwhile examining the change to the weights that the revised methodology lead to. Looking at Table 4, we see that most individual score weights have not changed from what they were in the Second Prototype. There was a significant decrease (33%) in the Peer Assessment score weight (before: 15, after:10), an even more significant decrease (40%) in the Document Assessment score weight (before:25, after:15) and a 50% decrease in the Self-Evaluation score (before:10, after: 5). There was also a small increase (16%) in the Offline Reputation score (before: 25, after: 30) which partially compensates for the lowered weights of the aforementioned criteria.

The changes may not have been dramatic, but they helped fine-tune the ranking results. The Self Evaluation and Peer Assessment scores were reduced because in the very early days of the platform users and experts closer to the Consortium were registered and endorsed each other; while the combination of those two criteria has a pleasant side-effect of encouraging engagement with the Platform and offering incentive to experts to spread the word about it to their colleagues, it would have been a strategic mistake to allow this effect to dominate the RMS ranking at an early stage when the biggest names in EU politics were still not yet activated users in EurActory. These criteria, being user-input criteria, will work very well when a large user base is there to provide appropriate input in large quantities. The same holds for the Document Assessment criterion which if weighted too high at an early stage would allow a single positive document evaluation to have a very significant effect on the reputation of the document's author; with many document evaluations the criterion would work well, with only a handful it does not give a reliable view of reputation so care should be taken not to have it dominate other criteria by giving it a weight that would be too high. The advantage of having multiple criteria is that they can complement each other so that one criterion covers for the shortcomings of another. By moderating the importance on criteria that relied on user input in the platform, we allowed criteria that did not (Business Card, Network Value, Offline Reputation)[9] to do their part and provide a convincing ranking which had a positive reflection on the Consortium's efforts and offered good publicity to the EU Community project.

As the uptake of the EU Community Platform continues to increase, it is expected that further weight adjustment will be carried out. Given the evolving nature of the RMS data, the fact that the revised weights adjustment methodology is both results-oriented and dynamic (in the sense that it can provide weights optimal for the time when it is applied) has been instrumental in the success of the RMS so far and will continue to be of great value in the future.

[9] We did not include the Past Measurements criterion in this list for two reasons: a. it does depend on user input (on the user input that played a role in determining the older reputation scores that the criterion takes as input) and b. because for the purposes of the event we effectively reset the system and deleted all previous scores (because they came from earlier versions of the RMS that had not undergone the quality assurance process the third prototype went through, weight adjustment included).

10 External Experts' Feedback

While in many respects the development of the RMS has followed the User Requirements (D2.3), in others it has evolved past them. At any rate, the WP3 Partners felt it was essential to validate the design of the RMS taking advantage of the diverse pool of external experts engaged in the workpackage.

WP3 External Experts were asked to review D3.2.4 "Reputation Management System (second version)" and respond to questions regarding the critical factors on the determination of a person's reputation and rate the relevant reputation criteria according to how significant they consider them. Their answers to these questions, obtained by means of web-based questionnaires, have been summarised below.

In a second round of expert feedback based on a near final version of this report, External Experts were asked to provide an assessment of the RMS along the lines of three open questions. Their answers are submitted verbatim alongside but not as part of the present report. The external experts raised a number of interesting issues that have been taken into consideration when preparing the present, final, version of the deliverable.

10.1 Perceived Importance of Individual Criteria

First of all, the eight reputation criteria adopted in the EU Community RMS were listed and briefly explained in order to be evaluated by the external users. A similar process was followed during the requirements gathering phase and resulted to the initial specifications for criteria weights in D2.3.

Peer-Assessment is considered as the most critical one, while Document Assessment and Business Card Reputation occupy the second and third place respectively. Past Measurements, Offline Reputation, Proximity Trust and Self-Evaluation have been rated as less critical, whereas Network Value is considered rather as a neutral factor.

Table 5: External Expert's Feedback on Criteria Importance

To what extent do you consider the following reputation criteria critical to determine the reputation of one person participating in the policy debate?	**Very Uncritical**	**Uncritical**	**Neutral**	**Critical**	**Very Critical**
Self-evaluation: The ability of users to define the areas of his/her expertise	0	22.2%	33.3%	22.2%	11.1%
Peer-assessment: The ability of users to endorse each other in different areas of expertise	0	0	0	55.6%	44.4%
Business Card Reputation: The reputation of one's person job position combining the reputation of the organisation	11.1%	0	11.1%	77.8%	0

To what extent do you consider the following reputation criteria critical to determine the reputation of one person participating in the policy debate?	Very Uncritical	Uncritical	Neutral	Critical	Very Critical
and his/her position in the organisation					
Document Assessment: The reputation of the documents a person has produced	0	0	33.3%	22.2%	44.4%
Proximity Trust: The level of connectedness between two persons as determined by social graph distances or organisational charts	0	22.2%	33.3%	22.2%	22.2%
Network Value: The influence of a person as the sum of the reputation of his/her network / connections	0	33.3%	22.2%	22.2%	11.1%
Past Measurements: A person's past reputation	0	0	44.4%	55.6%	0
Offline Reputation: The reputation of persons not necessarily captured in their online presence	0	0	55.6%	44.4%	0

It is interesting to compare the above results with the weights derived using the Revised Weight Adjustment Methodology ahead of the project's first official ranking unveiling (Energy Union at the European Parliament).

In some cases, the WP3 External Experts' opinions appear to be in line with decisions to change weights.

For instance, the Self Evaluation criterion which was deemed largely unimportant by the WP3 External Experts, initially had a higher weight (10) than the one later decided (5). The reason for the change was this: at a relatively early stage of adoption of EurActory, the experts that have signed up and self-assessed had an advantage over those that had not, so we decided, on the basis of domain expert evaluation of the results obtained, that we needed to tone down the effect of Self-Assessment at the present stage. The fact that this decision is in line with our External Expert's opinion is fortunate, but there are no guarantees that the two methodologies will always agree.

In other cases, following the Experts' opinions would have lead us to entirely different decisions than the ones we took.

For instance, the Peer Assessment Criterion which was deemed to be the most important criterion according to WP3 External Experts had a weight of 15 in the previous RMS prototype and this was reduced to 10 ahead of the event. The Network Value on the other hand which appears to be the least important criterion in the WP3 External Experts had a weight of 15, which remained unaffected prior to the event, making it a more important criterion for the RMS than the Peer Assessment Criterion. The low importance attributed to the

Network Value criterion was surprising and could indicate a certain aversion towards social media as a source of reputation. The fact of the matter is that if an expert X endorses an expert Y on a topic T within the platform, this is indeed an indication of Y's reputation. But if that same expert X who happens to be a top expert in topic T follows expert Y on Twitter, that is also, potentially, an indication of Y's reputation on topic T. Arguably, there is a degree of uncertainty involved because there is not an explicit act of endorsement on topic T. But if 90% percent of the top experts on topic T follow Y, that seems like a good enough indicator of reputation. It is numbers where the Network Value Criterion has the edge over the Peer Assessment Criterion. The Network Value Criterion is informed about any relation between experts X and Y with Twitter accounts, whereas the Peer Rating criterion may only record endorsements from users of the platform to other experts. Therefore, at this stage, the Network Value criterion offers a better view of experts' reputations. The Peer Assessment Criterion will become more and more useful as the number of users grows and more specifically as the number of endorsements collected grows.

What is obvious from our experience both with adjusting the criteria and asking for opinions about their importance is that such opinions may be useful in understanding how the criteria are perceived but cannot be the (main or only) basis for a methodology determining the weights. It has proven much more fruitful for the team involved with the RMS to ask domain experts for feedback on specific rankings and debug / adjust the RMS accordingly, for one simple reason: domain experts have an opinion of the reputation of individuals they know. The RMS is a complex system performing a multitude of complex computations in order to produce reputation scores and rankings. It is nearly impossible for external experts to take into consideration all pertinent factors and assign optimal importance to individual criteria. It has taken the creators of the RMS significant effort in debugging and experimentation to gain some degree of understanding of the effect of individual weights and changes in the usage of the EU Community platform may have in the rankings. It is non-optimal to leave this hard-earned experience unutilised and equally non-optimal to not ask domain experts questions they can provide useful answers to. The revised weights adjustment methodology takes these lessons to heart.

To conclude, the external experts' opinions on the importance of individual criteria are not a suitable basis for setting criteria weights. The methodology we have in place is much better suited for this task than the alternative of relying on experts' opinions on the importance of individual criteria. At the same time, the opinions that we have got from the WP3 external experts is useful in the sense that they demonstrate an attitude towards individual criteria. This is invaluable information, we simply cannot ignore as it relates to how the entire RMS is perceived outside the Consortium. Their feedback can guide us in better explaining the criteria as well as in giving us incentive to try different weight combinations in the future. More importantly it should serve as a guide for future developments of the RMS, in adding and re-designing criteria, as the RMS's credibility largely relies on having criteria that intuitively capture important aspect of reputation.

10.2 Opinions on Methodological Choices

WP3 experts were also asked to evaluation a number of methodological choices made in the RMS. Some of the choices up for evaluation involved the RMS as a whole and others concerned specific criteria.

The detailed results are presented below. It is worth noting, the strong support for the RMS's emphasis on topic-based reputation computations and the decision to base it on multiple criteria whose scores are combined as a weighted average, as well as the overall support for all decisions made in the design of the RMS experts were asked about.

Table 6: External Expert's Feedback on Methodological Choices

To what extent do you agree with the following methodological choices regarding the calculation of the reputation of one person participating in the policy debate?	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Define reputation on a per topic basis instead of an overall reputation score per user	0	22.2%	11.1%	22.2%	44.4%
Synthesise independent reputation criteria values on the basis of their weighted average	0	0	22.2%	66.7%	11.1%
Use social media as reputation data sources (e.g. LinkedIn, Twitter)	0	0	33.3%	55.6%	11.1%
Combine crawled reputation data with manually provided user's input	0	0	33.3%	33.3%	33.3%
Documents reputation score are calculated based on their quality, relevance with the policy process position endorsement	0	0	22.2%	33.3%	44.4%
In document assessment multiple author's reputation is assigned in full to all of them	0	11.1%	33.3%	55.6%	0
In peer assessment, the reputation of the endorser is taken into account	0	0	0	55.6%	44.4%
Peer-assessment between persons that work together have different weights	0	0	33.3%	22.2%	33.3%

10.3 Past Reputation

The WP3 External Experts were consulted also on the timeframe of the Past Measurement criterion. It seems that the more past data available the better (even dating more than 4 years ago), synthesized into a weighted average. In this phase, the time range of this data is limited due to the shorter life of the project. The feedback obtained, however, will be taken into account during the exploitation phase of the project tools.

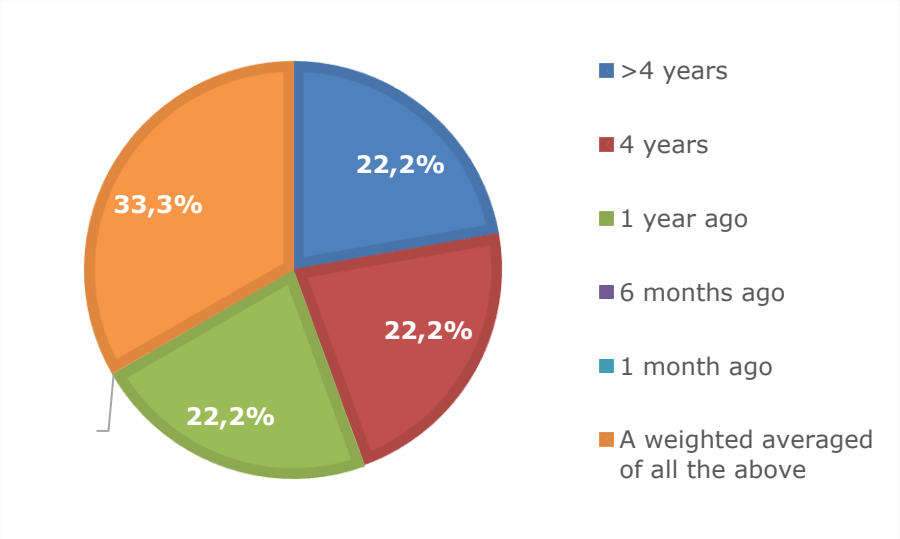

Figure 6: Answers to the question "Which time frame should be taken into account when calculating one's person past reputation?"

10.4 Offline Reputation

One of the most interesting aspects of the RMS is the inclusion of an Offline reputation criterion. This sets the EU Community RMS apart from comparable systems relying entirely on on-line information and equips the RMS with a tool of assessing the reputation of individuals who for one reason or another (age, choice, other) may not be picked up by the online radars of the RMS. It was deemed important to gather ideas about what kind of information should be taken into consideration.

In the open questions regarding the input that will be valuable to calculate the offline reputation of someone, experts have raised the issue of relying on facts and stated as examples, publication (scientific, articles, policy papers, comments) and speaking records, participation in Expert Groups, social connectivity, skills and job performance, surveys to peers or general public. Some of the above (e.g. surveys to peers and social connectivity) are partially covered by the current module, while the rest are out of the scope of the EU-Community project. However, they have been included to identify potential extensions of the EU-Community Reputation Management Module beyond the project and spot research challenges, in the field of Digital Reputation Management.

The question posed to the WP3 External Experts was:

> What input will be valuable to calculate the offline reputation of someone?

The answers obtained were:

- Job performance, social connectivity (social capital), publicity. social responsibility, training, skills, knowhow (intellectual capital)
- Publications, peer assessment
- Publication record. Speaking record. Expert Working Group participation
- Articles, policy papers

- Input from previous employer
- "Facts, facts and facts..." offline (as well as online) reputation must be based on facts (not on "self-evaluation!). The users should only give "facts" and not see the "immediate" effect on the reputation calculation (if so they will start manipulating). However, the method on how everything is calculated must be public and documented
- Results of ad hoc polls/surveys done to the purpose. Personal track record derived from public sources (including newspapers, official registries, etc.) Interviews with friends, colleagues, relations etc.
- Number of printed publications on a topic either from the person or co-authored by the person, reputation score of people who would write pre/postface or comments on printed publications, number of public conference keynotes of the person, online referencing of the offline reputation of the person by peers or by the general public

In addition to interesting, albeit not necessarily always practical, suggestions for offline reputation, one can easily see a number of suggestions for factors contributing to Offline reputation where there is obvious overlap with online reputation criteria. For instance, the answer mentioning "Articles, policy papers" may seem to be off the mark, given that there is a Documents criterion. However, an overlap of this kind is indeed justified since important documents authored by a policy expert may be associated with, for instance, older policy processes, not in PolicyLine. Offline reputation may indeed take into consideration authorship of important documents, (real-world) social relations etc. without undermining or coming into conflict with specialised online criteria.

11 Policy Domain Experts' Input Collection

Whereas WP3 External Experts provided feedback about the way the RMS was designed and implemented (Chapter 0) to the RMS technical team, policy domain experts provided input about the reputation of other experts in their domain of expertise.

The main three axes for that engagement were the three pilot topics:

- Energy Union
- Innovation and Entrepreneurship
- Future of the EU

11.1 Objectives

The RMS is an online tool providing reputation scores to key actors (experts) in specific policy domains. It only collects reputation information, computes reputation scores and ranks experts that it knows about, that is experts listed in EurActory. While the Crawlers Component can automatically discover experts and create EurActory profiles for them on EurActory, it was never the assumption that all experts will be automatically discoverable. So, one thing we need the policy domain experts we already knew to tell us is which policy experts are missing from EurActory, so that we can include them.

The other key thing we needed policy domain experts to let us know was their opinion on other experts so that the ranking for each topic would not be solely depend on the automated criteria scores of the RMS, but would result from blending those scores with the policy domain experts' opinions of their colleagues.

In summary, the main objectives of this process were:

1. To ensure the top experts on each policy topic are included in EurActory
2. To ensure sufficient quantities of policy domain expert input are provided for the ranking to be as reliable as possible

11.2 Methodology

To best achieve the two objectives for the task of bringing in policy domain expertise into the RMS, we took a multi-step iterative approach with an interesting feedback loop.

Had we, somehow, started with a group of experts E_0 and asked them who else is to be included and to evaluate themselves and their peers, we would have got the kind of information we wanted, but unless the initial group of respondents was particularly large and representative we would have been faced both with problems of quantity and unrepresentative bias.

Both of these problems would be somewhat minimised if we were to also take into consideration not only input from the initial set of experts E_0 but from a set E_1 consisting of both the initial experts and the peers they have named as noteworthy experts in the same policy domain. This, of course, assumes that experts have a broad knowledge of the policy domain and know and respect peers from different countries than their own, of different backgrounds and with different roles and interests in policy processes – an assumption we are indeed making.

On this basis, it is obviously preferable to repeat the process as many times as feasible given the time and effort constraints of the project, expanding the set of experts providing input from E_0 to E_1 to E_2, … to E_N.

It should be clear that:

1. Given enough iterations (*F)* this process would eventually reach a point where E_F and E_{F+1} are the same i.e. when the experts in E_F know of no additional experts not in E_F. After that there would be no point repeating the process.

2. It may not always be possible to reach that point due to resource constraints i.e. *N* will not necessarily reach *F.*

3. Even if it were the case that *N=F,* it is not guaranteed that E_F is the set of all noteworthy policy domain experts (let us call this set *E*). Even assuming experts will do their best to name other noteworthy experts, they may simply forget or may simply not know the experts that have not yet been included. (If every expert knew and could name every other expert in that policy domain, we would have needed only one expert to let us know who is who in that policy domain.)

In other words, our process comes with no guarantees that it will produce the perfect outcome. What it does do is allow us to have a degree of reassurance that given a reasonable, but not complete, not necessarily unbiased initial set of experts we will be able to expand the set of respondents with a reasonable amount of effort, to achieve a reasonable, fairly complete, and fairly unbiased, outcome.

Whereas it would seem that the process for ranking the list of experts could follow the process of expert discovery, in fact the two mostly went hand-in hand.

The bulk of the interaction with experts happened in small expert group workshops (with approximately 20 experts) where experts were asked to both name noteworthy peers and to name the top 20 most influential individuals per topic. An offline reputation ranking was created on the basis of mentions in the answers to the top 20 experts question which was then converted into an offline reputation score on the basis of the most mentioned experts receiving 100 points, the least mentioned experts receiving 50 points, and the remaining experts being given a score in between on the basis of their off-line ranking (e.g. if there were 51 experts each with a different number of mentions, the top ranked expert would get 100 points, the second highest rank expert would get 99, …, the expert mentioned the least number of times would get 50 points and experts not mentioned in any experts workshop would get 0 offline reputation points).

There were also one-on-one sessions with experts where the format of the meetings was less structured but nevertheless involved both questions about who must be included in the ranking and who the top experts are. In this instance, experts were asked to peer rate whomever they thought to be a top expert in one or more topics.

11.3 Sustainability

The process described has two main aims and two corresponding components: expert discovery and expert rating. Both have been supported by the Community Manager as part of the pilots. This may raise concerns about how it is to be continued after that period.

The answer to these concerns is twofold:

- The approach was tested during the project pilots exactly under the same conditions as it is expected to function is a production setting. Intensive support from the Community Manager will only be needed if/when a new topic is added to the platform. As was the case in the pilots, the Community Manager's role will be to ensure enough policy domain expertise on that topic is gathered and made available to the RMS before the ranking on that topic is first unveiled. This is a measure that will ensure a high standard of quality for new rankings. The Community Manager's intense involvement is part of a bootstrapping process that allows both the per topic community and the respective ranking to mature quickly and in an organised manner before the first official EurActory ranking for the topic is unveiled.

- In a production setting, after the initial ranking is first unveiled, it is expected that significant amounts of reputation-related information will be coming to the platform organically through its gradually expanding user base. In addition, as the platform gains traction and the expert community finds out about it, it is expected that new users will come in and create their own profiles. Again this is what we are testing out currently in the pilots and more specifically in the Energy Union pilot, the first for which an official EurActory ranking has been unveiled.

There may need to be further refinements (possibly more than one Community Managers spread across different locations within the EU, more streamlined online ways for users to invite their peers to EurActory, etc.), but the core ideas are sound. The RMS is designed to be self-sustaining. The additional help a Community Manager can bring when expanding to new policy domains, allows reaching a point where there is no compromise of the quality of displayed results before relying on the RMS's self-sustainability. This is a pragmatic solution to a real problem that will appear in a production setting, tested out successfully during the pilots phase.

11.4 Results, Analysis, and Conclusions

At the time of writing, efforts for policy domain experts input collection and for finalising the rankings for Entrepreneurship and Innovation and the Future of EU topics are nearly complete, and so is an Energy Efficiency ranking and a

country-based Energy Union ranking. Here we will focus on the Energy Union Top 40 ranking unveiled in February 2016 at the European Parliament (Figure 7).

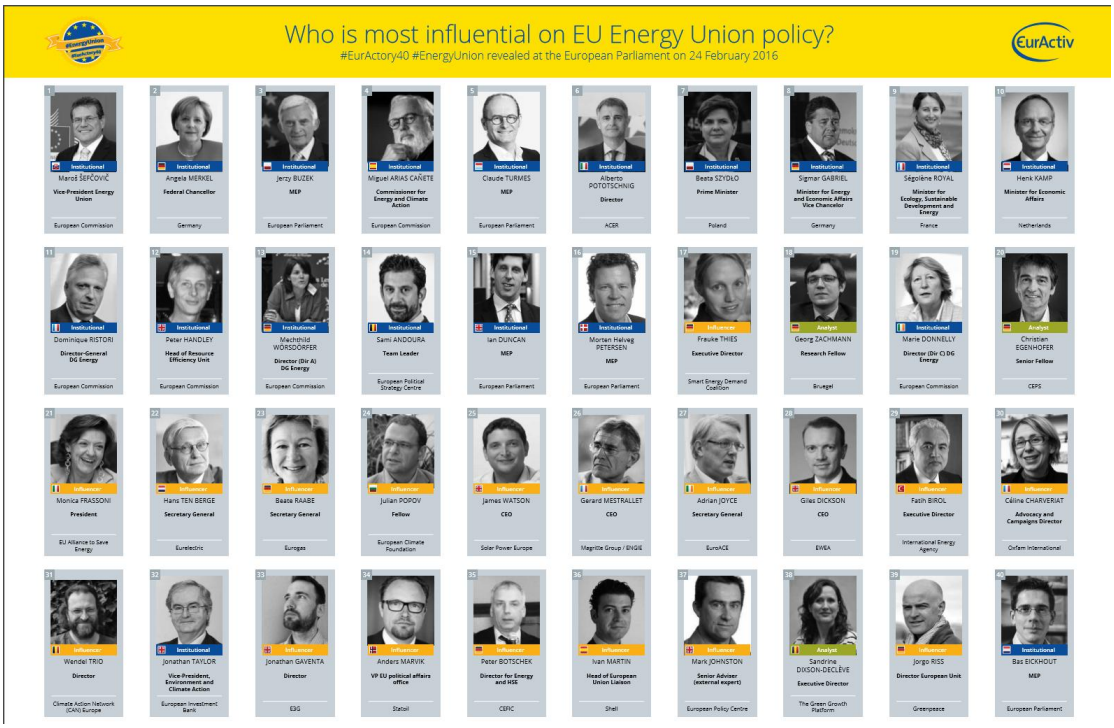

Figure 7: EurActory Energy Union Top 40 Ranking on 24 February 2016

The EurActory Energy Union Top 40 Ranking was very well received and this is a measure of the success of both the technical and the policy domain expert input collection efforts that lead to it.

But there are more things to be said about the top 40 ranking obtained. (The Energy Union ranking is not limited to 40 positions, but the top 40 as this is in itself an interesting basis for discussion and conclusions.)

The infographic summarising some of the key facts about the top 40 Energy Union ranking (Figure 8) prepared for and presented at the event, reveals some interesting statistics.

The majority of top influencers were in Brussels (87.5%), while things were much more balanced in the case of EU/national political leaders and EU/national government employees (EU:55%, national:45%) and the handful of analysts that made it in the top 40 were all based outside Brussels. On this front, one could note that despite the fact that the project was EU-centric and as a result Brussels-centric, there was a welcome balance national and EU players. At the same time there was an imbalance between genders, albeit this was at the typical levels witnessed in the top echelons of EU policy involvement, therefore a problem of the realities of EU policy not of the top 40 ranking which merely reflects them.

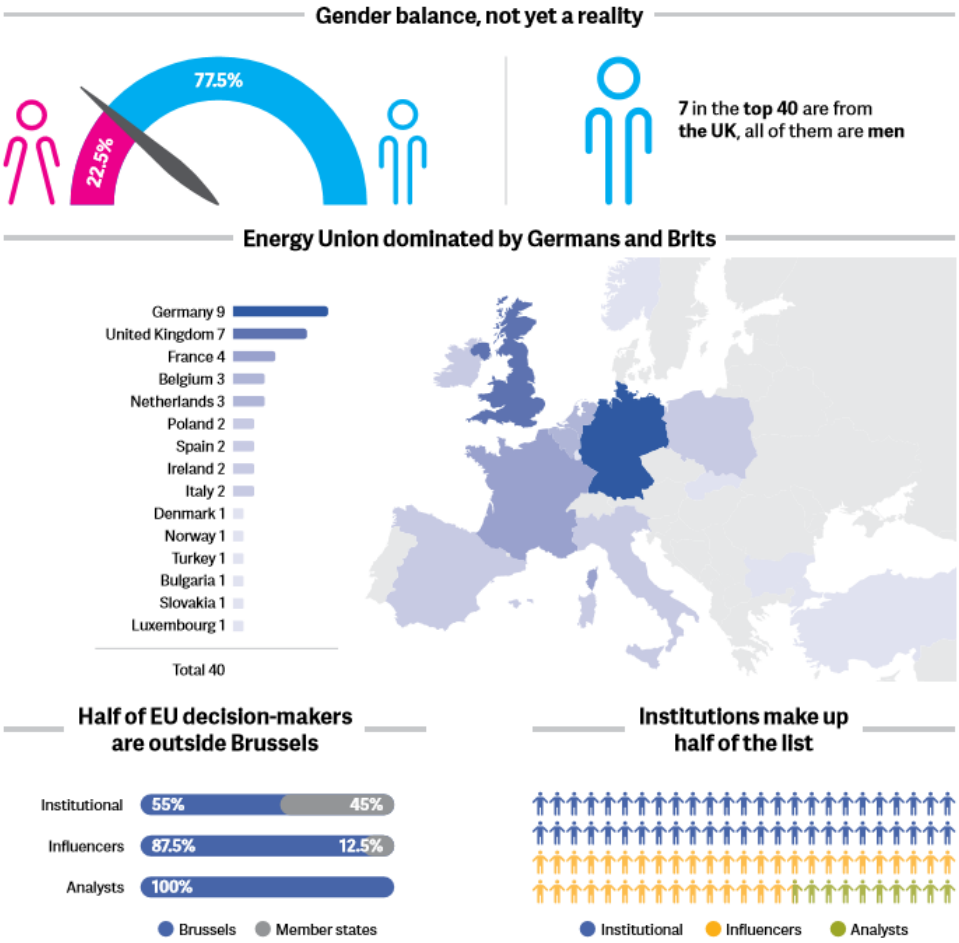

Figure 8: Interesting facts about the EurActory Energy Union Top 40 Ranking of 24 February 2016

The fact that exactly half the top 40 experts were elected political decision makers or held significant positions in EU institutions, followed by a large number of influencers and a very small number of analysts, shows that "reputation" has been taken to be roughly equivalent with "ability to influence EU policy".

As stated in the introductory chapters there is nothing wrong with that. The nature of the "reputation" a reputation system computes is what it is and comes about as a result of design decisions and user input, in this case, especially the latter. This notion of reputation is both interesting and useful for the purposes of attracting attention to the EU Community project.

The notion of scientific expertise is also an interesting and useful one. It would have been possible for the RMS to be computing this notion of reputation instead, if it was primed with input from academics and researchers.

This is an important point. We do recognise that the experts we engaged have caused a particular bias in the ranking and shaped its nature. However, both the selection and the bias are in line with the dissemination and commercialisation efforts for the EU Community platform as the target audience is individuals involved in policy making rather than a research/academic audience.

What we have come to understand and accept is that by having a single notion of reputation, this would have to be one that is most relevant to our target audience at the cost of providing an alternative one that would also have great merit and interest. We have more to say on the matter in the Conclusions and Future work chapter.

To recap, the results were influenced (by means of input provided) and meant to be positively received by the users of the EU Community Platform. On the basis of their reception, as well as upon critical examination of the end result, we conclude the effort has been successful in meeting its objectives. At the same time, we have come to understand that while our efforts centred around a specific and appropriate notion of reputation, one or more alternative notions of reputation are being ignored and this is an issue that, given our current understanding, could be potentially be addressed in future work.

12 RMS Data and the Issue of Privacy

12.1 Private Data in Reputation Calculations

The EU Community Platform is primarily an open-data platform. Specific types of private data are collected for specific purposes, and used for these purposes exclusively. Private data are protected by the platform and are not shared with third parties, except as required by law (though due to the nature of the EU Community's private data, the likelihood of there ever being a legal basis for a subpoena for those data is practically zero).

There are three types of private data used in reputation scoring:[10]

1. Peer evaluations (expertise endorsements) in EurActory used to compute the Peer Assessment criterion.

2. Answers to quantitative evaluation questions in Document Assessment questionnaires in PolicyLine, which are used both for computing a document prominence score used within PolicyLine and for computing the Document Criterion score used in both PolicyLine and EurActory.

3. Outcome and Duration Predictions for Legislative Procedures in Policy Processes in PolicyLine, used to compute the Predictions Accuracy criterion.

12.2 Public Data in Reputation Calculations

The majority of Reputation Calculations rely exclusively on public data:

1. Self-Evaluation: The areas in which a user declares he/she is an expert in are displayed publically in EurActory.

2. Business Card Criterion: The job position title and organisation where an expert is employed are public data displayed on EurActory (as well as other public sources, such as the Europa Who Is Who site and the Transparency Register, public Linkedin profiles etc).

3. Proximity Trust and Network Value: The information for the calculation of these criteria (Twitter 'follows'/'is-followed-by' relations and organisation where experts are employed) is also publically available.

12.3 Platform Data in Reputation Calculations

There is a class of data used in reputation calculations that falls under neither in the private nor in the public data domain; these are proprietary data that exist in the platform and are neither publically displayed, nor sourced from a public data repository (which means they are not public data), nor are they sourced by any user via a channel that identified the user as their source (which together with the fact they are not public, would make them private data).

[10] If Linkedin's Public API still supported obtaining a Linkedin user's 1st degree connections as was the case when the first version of the RMS was designed, these connections would be used in computing both the Proximity and the Network criteria and would be treated as private data.

1. The Offline Reputation relies on scores computed off-line by the Community Managers. The basis of these scores may change over time. At present, these scores are primarily based on input from the project's paid experts. In the future, it is envisaged that means such as collecting input from events where participants anonymously identify key people they would have liked to see participate in future events would be utilised for the computation of this score.

2. The Past Measurements criterion relies on old reputation scores, which have been computed and recorded by the platform but are not made publically available.

12.4 Privacy Concerns

A large number of expert comments received about the Reputation Management Module have been concerned with the issue of `privacy'. Somewhat surprisingly, they are mostly unrelated to the issues pertaining to visibility or use of private data.

Instead, most comments centred around the issue of data control. A complete answer to this question addressing control over all data in the Platform is beyond the scope of the present deliverable. However, the answer to the question of what control a user has over RMS-related information (public, private, and platform data) is answered below:

1. The user may at any time change or delete the Self Assessment (public), and Peer Assessment (private) data they have provided. They may also change, but currently cannot delete, the Document Assessment (private) data they have provided.

2. The user may at any time update their job title and organization in Linkedin, which is one source where the relevant information may be obtained from. If the source of this information is Europa Who-Is-Who, they cannot update the information themselves, but if they move to a different position (or stop working for the EU organisations Who-Is-Who includes) the relevant information will be automatically updated.

3. The user does not have any direct control over the relation other users choose to have with them on social media (see Network and Proximity Trust criteria), nor on the Offline Reputation criterion data. Finally, a user does not have any control of their past scores used to compute the Past Measurements criterion.

Then there was also the question of whether users would have control of the visibility of the RMS score or ranking. The Consortium decision was that:

- There would not be any per-account tinkering of RMS score visibility; this was not something we wanted users to turn on and off as they pleased, but rather an aspect of the system where we would have a uniform approach applicable to all profiles

- The criteria and the information used in their calculation would be explained, but the details of individual and overall scores would not be made public. Overall per-topic scores are used to rank experts in EurActory and in weighting the expert's feedback on documents in

PolicyLine, but neither use of the overall score exposes it or the constituent criteria scores to users.

13 Conclusions & Future Work

13.1 First Prototype: A Solid Technical Foundation

The primary aim of the first version of the Reputation Management System was to provide a solid general design and architecture for the RMS.

Both those two key objectives were met.

The overall design of the RMS as a system with multiple criteria has proven extremely flexible and expandable. It has proven particularly easy to both change the implementations of individual criteria and to add new ones. The mechanism for combining the criteria scores by assigning weights and computing a weighted average has also been very convenient for adjusting the RMS.

The architectural decision to implement the RMS within CurActory as a set of stored procedures and functions is optimal both in terms of system performance and of integration effort, as:

1. it eliminates the need to have primary data transferred to the RMS (from CurActory and/or directly from the Crawlers Component and the Opinion Mining Component) and reputation scores sent back to CurActory (from the RMS)

2. it eliminates the need to provide and optimise the respective interoperability mechanisms

3. it eliminates the need of other components such as EurActory and PolicyLine to be gathering information from an additional component.

At the same time, at the level of data representation and data processing specifications, D3.2.1 outlined the solution implemented in all three RMS prototypes that allows two opposing requirements to be satisfied, namely the requirement that users not notice any delays due to reputation calculations (which is possible if reputation scores are computed off-line) and the requirement that users be able to customise the reputation calculation (which is not possible if reputation scores are computed off-line). The key two ingredients of the solution are: (i) a careful design of reputation scoring customisability and (ii) a division of labour in reputation scoring between an off-line pre-computation stage and an on-line light-weight computation stage utilising both reputation data prepared off-line and a user's reputation scoring preferences.

13.2 Second Prototype: Meeting and Exceeding Requirements

What the second version of the RMS brought to the table was conformance to the requirements in D2.4. It did so by

- adding two criteria (Peer Assessment and Offline Reputation) and
- providing better implementations for existing criteria:

- the Position and Organisation criteria, were both made topic-sensitive and joined into a single Business Card Reputation criterion and
- the Document criterion was re-designed to be able to work with PolicyLine, the main user-facing component of the EU Community platform, now that its first release is eminent.

However, it is not only the case that, in its second version, the RMS prototype no longer needed to play catch up with requirements in D2.4; in fact, it went beyond those requirements:

- It combined the Organisation and Position Reputation Criteria in the Business Card criterion (solving the problem outlined in D3.2.1 that having them separate could cause)
- It had a Network Value criterion (Criterion F) not foreseen in the user requirements
- It integrated the Past Measurements score as a first-class criterion score, which could be treated in a uniform manner to the other criteria.

13.3 The Third Prototype: Production-Readiness

The main aim for the third prototype was to solidify the existing work in the second prototype, subjecting the code to extensive testing, debugging and making improvements to it. This aim was achieved in the First Release of the Third Prototype presented on February 24, 2016 in the European Parliament. Whereas up to then the main concern of the RMS team had been technical correctness, the approach taken in the third prototype's Quality Assurance process was results-oriented and the bar was set much higher. The weights methodology was adjusted and in fact merged with policy domain expert evaluation of the results so that a feedback loop would enable continuous improvement of the weights were needed and/or of the algorithms and data (including clean-up of old reputation scores and careful gradual re-activation of the Past Measurements criterion four past data collection timeframes) where appropriate. More importantly still, during the pilots phase significant effort was invested in the collection of experts' opinions about their peers and these played a major role in shaping the ranking, turning it from a purely automatic one to a ranking blending automated criteria scores with the results of policy domain experts input.

From a purely commercial-minded point of view, the true test of the success of the RMS and the efforts to provide it with policy domain experts feedback would be its reception. The event at the European Parliament where the Energy Union ranking was unveiled was judgment day for the sum of our efforts. Reception of the event and the ranking results has been overwhelmingly positive, as gaged both during the event itself and in social media mentions (Figure 9).

Figure 9: Sample social media reactions about the EurActory Energy Union Top 40 Ranking event

13.4 Next Steps

There are two broad directions for the RMS, corresponding to its dual-nature as both a research project and as commercial product (or rather the basis of a commercial service).

From a commercialisation point of view, the decisions that have been taken so far regarding the visibility of information and the modalities for input collection are in line with relevant plans. It is the Consortium-wide wish that efforts to commercially exploit the RMS and other technologies bear fruit and that the outcomes of this project endure beyond its duration in a commercial setting.

Further developments on the front of criteria will be relevant on both the commercial and the research front.

Additional criteria aside, there seems to be a more fundamental potential future development that could provide a solution to the problem we faced in choosing between a notion of reputation that translated to "ability to influence EU policy" and a notion of reputation that translated to "knowledge (scientific or other) about the subject matter". As mentioned in Section 11.4, it has now become apparent that having a single reputation metric forces alternative notions of reputation to compete for dominance in a way that is counterproductive and unintuitive. It has also become apparent that this is not merely a matter of selecting the right criteria but about making the distinction from the outset in all aspects of the system including the collection of self-evaluation, peer assessment and offline reputation data. That said, there may be a need for additional criteria also. For instance, while the Documents criterion can be used to elevate the reputation of researchers and academics when their work is referenced in relation to a specific policy process, it may also be beneficial to have an automatic criterion looking at bibliographic referencing of scientific publications irrespective of their association with policy processes.

The issue of generic vs topic-specific reputation has also been of interest to both the reviewers and to the RMS technical team; whereas the DoW, the user requirements and also the External Experts opinions are in line with the RMS's current emphasis on per topic reputation, there is scope for further work in this area. The current deliverable describes a way of creating a generic reputation score out of pert-topic scores (Section 7.2), which is not biased towards specialisation as the previously used algorithm was, but other alternatives could be explored.

While we have tried to speak separately about the future of the RMS as part of a commercial service as opposed to a research project outcome, the truth of the matter is that both directions will be pursued; further research on the basis of a real-world evaluation of the RMS has as a prerequisite its commercial success and the commercial spin-off of the RMS will certainly stand to benefit from any further research whether it be an investigation into the links between reputation aspects and criteria, a way for automatically adjusting criterion weights or some other functionality making the RMS adopt automatically to data patterns (e.g. sudden group endorsements of each other) that could affect its reliability, and investigations along the recommendations of recommendations of our external experts such as the addition of one or more criteria that would work with unstructured crawled data or the development of a version of the RMS based on stochastic methods and a comparative evaluation against the deterministic RMS defined here. In fact, we believe that a long enough period of successful commercial operation will generate additional interesting real-world research questions to be answered.