Building a National E-Service using *Sentire*

Experience Report on the Use of Sentire: A Volere-Based Requirements Framework Driven by Calibrated Personas and Simulated User Feedback

Chris Porter
Dept. of Computer Science
University College London
London, United Kingdom
c.porter@cs.ucl.ac.uk

Emmanuel Letier
Dept. of Computer Science
University College London
London, United Kingdom
e.letier@cs.ucl.ac.uk

M. Angela Sasse
Dept. of Computer Science
University College London
London, United Kingdom
a.sasse@cs.ucl.ac.uk

***Abstract*—User experience (UX) is difficult to quantify and thus more challenging to require and guarantee. It is also difficult to gauge the potential impact on users' lived experience, especially at the earlier stages of the development life cycle, particularly before hi fidelity prototypes are developed. We believe that the enrolment process is a major hurdle for e-government service adoption and badly designed processes might result in negative repercussions for both the policy maker and the different user groups involved; non-adoption and resentment are two risks that may result in low return on investment (ROI), lost political goodwill and ultimately a negative lived experience for citizens. Identity assurance requirements need to balance out the real value of the assets being secured (risk) with the user groups' acceptance thresholds (based on a continuous cost-benefit exercise factoring in cognitive and physical workload). Sentire is a persona-centric requirements framework built on and extending the Volere requirements process with UX-analytics, reusable user behavioural models and simulated user feedback through calibrated personas. In this paper we present a story on how Sentire was adopted in the development of a national public-facing e-service. Daily journaling was used throughout the project and a custom built cloud-based CASE tool was used to manage the whole process. This paper outlines our experiences and lessons learnt.**

***Index Terms*—Requirements engineering for user experience, usability, enrolment process, Personas, Sentire, Volere, industry and research collaboration**

I. Introduction

A number of user centric tools, techniques and frameworks have been proposed for the specification, design and development of e-government services, however such services "*continue to demonstrate poor usability*" while providing a "*less-than-satisfying UX*" [1]. ISO (International Standards Organization) defines UX as "*a person's perceptions and responses that result from the use or anticipated use of a product, system or service"* [21]. We argue that this is an explicit result of a number of conflicting and competing goals, arising from governments, citizens and contractors alike where reconciliation requires an objective effort where each party is given a voice from the earlier stages of a system's lifecycle. Although usability and UX are seeing their way in government projects, a study conducted across 38 project contracts for public software systems [2] Lehtonen et al noticed that in most cases statements about usability were vague and/or focused on precise design or process features (e.g. *software must be tested for usability* and *screen must have an exit button*). Typically policy makers also specified that to ensure usability user representatives shall give a subjective rating following system demonstrations given by the contractors. Jokela and Buie [1] argue that these are not proper usability requirements since they lack verifiability, validity and comprehensiveness. Subsequently these are comparable to wish lists or strategic desires that may or may not support the user's primary task. We believe that a unified and systematic government-wide requirements framework that puts the user at the centre of the equation is missing. This must also embrace knowledge accumulation and re-use and thus reducing dependence on contractors' knowledge and experience. Such dependence can pose serious risks to project success [1]. Janowski, Estevez and Ojo [3] state that the lack of methodologies and models, as well as the lack of cohesion between related e-government projects are two of the major causes of e-government project failure. User eXperience is very difficult to quantify and thus specify in measurable terms within the requirements document or call for tenders. For this reason UX may find itself on the back-burner and contractors may de-prioritize it so as to increase competitiveness in terms of costs and time to deliver [1].

Through our Volere-inspired requirements elaboration and design framework – Sentire – we are studying ways to express the impact that enrolment-related e-service requirements can have on users (at requirements stage) in terms of perceived workload and willingness to adopt. We decided to focus on the design of acceptable enrolment processes since we believe that this is a critical success factor within the e-government domain. France [18], UK [17, 19] and Austria [20] are some of the countries that have experienced negative repercussions due to sub-optimal identity related processes. One of our main goals is to provide a framework and corresponding tools that would help designers and service providers (in this case government entities) design better e-services. We define the term 'better e-services' as systems that are human-centric and conducive to goal achievement, both from a citizens' perspective, but also from a service provider's point of view. This framework positions users at the centre of the requirements process and adopts calibrated personas (classical personas embellished with re-usable statistical behavioural models) to generate simulated

user feedback on critical design decisions (e.g. enrolment processes). This simulated user feedback acts as an immediate and objective starting point for the design team to assess the impact that certain design requirements might have on various user groups – enabling an iterative process of progressive improvement. This mitigates against bad design decisions being identified at a later stage (prototyping) making it more difficult and expensive to fix, or at worst, not being captured at all during user acceptance testing. The persona calibration process is based on a systematic process of understanding which also provides interesting insights on our assumption personas thus informing persona development and evolution. Sentire was used in previous case studies and interesting insights emerged [7, 10, 13]. We confirmed the mechanics in the earlier studies and validated results in subsequent ones. However we wanted to assess the latest iteration of Sentire within a fully-fledged e-government service project, from inception to launch. This will allow us to focus on the overall process and determine the impact that Sentire affords at each stage.

This paper is structured as follows. Section II provides some background on the theory behind this work (Volere and Sentire respectively) as well as project specific information including our main collaborator, the e-service to be developed, our research goals and techniques adopted. We then present the process we followed to build this e-service in Section III including our own reflections at each stage. Finally we present our conclusions and recommendations in Section IV, followed by some future work in Section V.

A. Companion Resources

This document is as self-explanatory as possible however the following resources offer a gentle introduction to the central techniques discussed throughout.

- Sentire poster: *http://goo.gl/Giy4ar*
- Sentire explainer video (basic): *http://goo.gl/9OjPKM*
- Explainer video (advanced): *http://goo.gl/x4pACJ*

II. BACKGROUND

A. What is Volere?

Volere is considered to be a practical requirements framework recommending usable notation, techniques and deliverables within a systematic and rigorous requirements elaboration process. This can be contrasted with the more formal requirements elaboration techniques, such as *i**, which may be too complex for non-technical stakeholders to follow and comprehend [14]. Volere encapsulates industry best practices for requirements elaboration and specification encouraging clarity, explicit scoping, simplicity, rigour (through specific milestones and templated deliverables), scalability, traceability, validation and testability (adopting *Fit Criteria*), re-use and a socio-technical perspective within a repeatable process [4, 14]. A complete discussion on Volere is provided in [4].

From our experience, Volere has been received positively by both policy makers and IT specialists, specifically for its rigour, process clarity and simplicity.

B. What is Sentire?

The name Sentire ("to listen") is inspired from its underlying process Volere ("to want") and reflects the idea of listening to end users through simulated feedback generated computationally using statistical user behavioural models: *calibrated personas*. This feedback is generated at the earliest stages of the requirements/design process. This technique does not replace traditional UX evaluation techniques (e.g. user walk-throughs, focus groups and eye-tracking studies) but provides an early-stage user-centred discipline to inform decision making on critical design aspects. Generally actual users are introduced to the process at a later stage, especially when a hi-fidelity prototype is developed. Sentire bridges this gap, introducing the users' reactions to critical design aspects, potentially from day one. This a) provides a low-cost and immediate feedback mechanism, b) reduces the risk of major re-work on critical aspects and c) instils systematic user-centricity as part of the requirements and design process where every decision taken will result in some form of impact on the users' experience, represented as a vector – containing the magnitude of such impact and also its direction (negative vs. positive). In line with Nielsen's [5] task success rate measurements, Sentire provides indications of the probability that a given user group would be willing to complete the task (success). At the same time Sentire also provides NASA-TLX-based workload measurements [6] as well as other metrics acting as meta-information to explain the given indications (adopting NASA-TLX's multi-dimensional nature). Sentire allows policy makers to specify UX related requirements such as task success rates (e.g. 80% should be able, and willing, to enrol on our service with a 95% confidence interval) and workload (e.g. perceived workload should be kept below 40% for all user groups). In collaboration with the contractor, the project team would then work towards reaching the desired goals, from day one. Sentire extends the Volere process as shown in Figure 1. In Section III we will be explaining each extension as well as their practical implications.

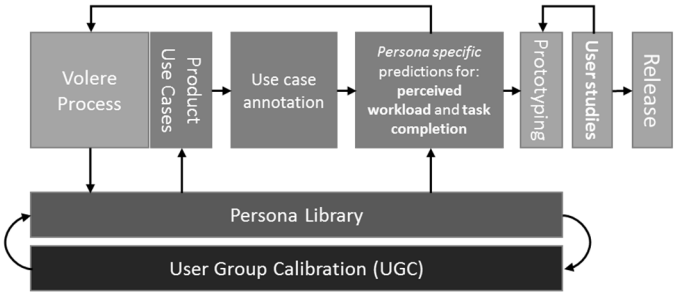

Fig. 1. Sentire requirements and design framework – extending the Volere process to enable user feedback simulations at the requirements stage

The persona library holds statistical user models that explain user behaviour when facing critical design factors. For this study we are modelling perceived workload and willingness to complete a task when faced with different enrolment processes. These user models are based on a persona calibration exercise which captures user behaviour while performing a series of pre-determined tasks. This information is then processed in a statistical package to fit two regression models (linear regression model for workload data and binary logistic regression model to

predict the users' willingness to adopt the service, explained by the probability that a user will complete the enrolment process). These models provide regression coefficients which explain the user group's attitudes towards the different design factors behind enrolment processes. These coefficients could then be used to understand the potential impact of different design alternatives on users (in terms of willingness to adopt the service and perceived workload). The tasks presented during calibration are configured differently based on design elements identified in a previous empirical exercise [7, 10] (i.e. *number of fields to fill*, *possible delay severity*, *interruptions to daily routines*, *number of new credentials* as well as *level of service compulsion and frequency of use*). User models are then plugged into the respective traditional persona construct representing similar users as the ones used to generate the user models – termed as *calibrated personas*. As a case in point, a behavioural model for undergraduate students can be re-used with various personas sharing the same set of demographics. Product use cases (i.e. those involving an enrolment process) are annotated with values representing enrolment specific design elements, and are also associated with active actors (represented by calibrated personas). This allows the design team to generate user experience simulations at the requirements stages, before any prototypes are built.

In order to facilitate this otherwise laborious process, a CASE tool was developed to assist designers in the aforementioned steps. The tool facilitates the generation of deliverables compliant with the Volere requirements templates. This provides an overarching structure which is highly desirable especially at the initial stages of the development lifecycle. The CASE tool is also supplemented with an online calibration portal. These techniques and tools will be discussed in Section III, however more in-depth information is provided in [7, 10, 13].

The following tools and techniques are typically adopted in Sentire-driven projects: *Sentire's CASE tool* [7, 13], *calibrated personas* [10] which in turn build upon Cooper's *personas* [8] and Faily and Fléchais' *persona cases* [9], *participatory design* [12], *user walkthroughs*, *card sorting* [11] and *eye gazing tests* [16]. All of this is organized within *Volere's requirements process* [4], including its templates and deliverables.

C. The E-Service

The Malta Competition and Consumer Affairs Authority (MCCAA) are aiming to improve the way citizens interact with the authority while streamlining internal processes for increased efficiency. For this reason a series of meetings were carried out with top management to discuss their needs as well as our objectives, and eventually a collaborative agreement was formalised for the development of the Consumer Advice Portal (CAP) - a public facing e-services that acts as the first point of contact for consumers; offering an advice and complaints wizard, a publication repository, frequently asked questions as well as a knowledgebase on past cases. This e-service will also act as an internal knowledgebase enabling easier access to information and improve knowledge transfer and re-use for current and future case-officers. Based on this we agreed to adopt Sentire to develop CAP's requirements as well as design and build the actual portal using an agile development methodology (categorised as a *Rabbit* project in Volere terminology).

D. Our Goals

We believe that the enrolment process is the first hurdle in any e-government project and it can make or break a service. Users may opt not to use a service simply because the workload involved to enrol exceeds their acceptability thresholds. Sentire was specifically built to inform the design process through simulated user reactions given specific service alternatives. This case study allows us to reflect on and evaluate Sentire and its associated toolset within a fully-fledged national e-government project.

III. The Process

Figure 2 outlines the workflow adopted for this project, broken down by work unit. Each work unit will be tackled in the following discussion, highlighting implications arising from the adoption of Sentire.

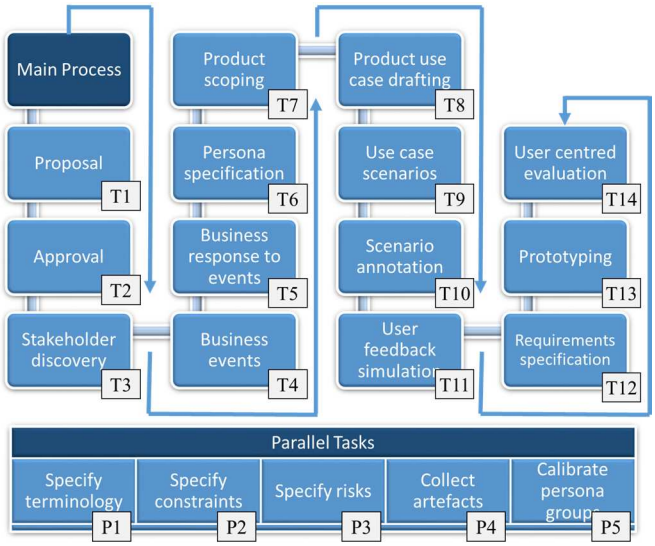

Fig. 2. Consumer Advice Portal project workflow. *T* represents sequential tasks while *P* indicates parallel tasks

In the initial stages (**T1, T2**) high level goals were specified and approved by top-management. A project team was also assigned. At this stage we started discussing entities that may have a direct or indirect influence on the e-service (**T3**). For this purpose we adopted a simplified version of Ian Alexander's onion-ring model for stakeholder identification [15]. Several pre-determined categories (from the Volere template) were also used to inform the discovery of potential people/entities/systems that can have an impact or are impacted by the authority's work. By the end of the requirements process around 47 stakeholders were discovered together with their inter-relationships (where applicable) and these included domain experts, internal specialists and other government entities.

A discussion on business events was initiated (**T4**). These are events that occur at the authority, including externally initiated events (e.g. consumer calls the help-desk) and time-triggered

events (e.g. bi-weekly batch processing of complaints). We feel that by creating an outline of business events we would be preparing a solid foundation to a) rigorously identify business use cases (how the authority responds to business events) and b) discover new areas that may need further investigation. These are foundational steps enabling a fine grained scoping exercise for the system-to-be. By the end of the requirements process 12 externally initiated events were identified along with six time triggered events (18 in all). At this point we started to understand how the authority operates and how different internal stakeholders respond to both external and time triggered events. This helped in formulating a first set of technology-agnostic business use cases (**T5**) explaining the various work processes and exceptions. Relevant artefacts, such as physical forms, documents, leaflets, meeting recordings, transcripts, photos and correspondence were stored within the project's workspace in Sentire's CASE tool (**P4**). So far all discussions revolved around the authority's work and there was no mention of any e-service related features. Initially we wanted to draw a picture of the authority's world and work as well as any form of interaction with consumers and traders. This provides more knowledge and data into which we can ground well-scoped decisions, rather than merely coming up with requirements for unknown problem/s existing in an undefined domain. Scoping becomes much easier and robust when based on broad knowledge. At this point, workflows started to emerge and various (sometimes unexpected) domain experts were involved in order to walk us through and explain the various scenarios. Business use cases were specified within the CASE tool reflecting how the authority responds to the various events. Information held included use case pre-conditions, business rules and outcomes as well as normal, alternative, exception, misuse and negative use case scenarios. Active and interested stakeholders were associated with use cases. Several domain specific terms and acronyms started to emerge and these were recorded in the terminology library within Sentire's CASE tool (**P1**). The terminology library facilitates communication across the entire team (including future team members) by providing a common dictionary of domain and project specific terms. Project constraints (**P2**) and risks (**P3**) were discussed in tasks **T1** and **T2** and also stored in their respective repositories. Furthermore, identified risks can also be linked to other aspects of the project (e.g. Use Cases) that can help to mitigate their occurrence. An online, centralized and structured repository makes it easier for team-members to monitor progress without the need to request and go through any ancillary documentation.

Following the exploration of business events and associated responses we have now formulated a clear idea of the main citizens groups who interact with the authority. A set of assumption personas was outlined while flagging those who might eventually make use of the consumer advice portal (**T6**). At this point a parallel task was spawned – persona elaboration, grounding and calibration (**P5**). Assumption personas were stored in the CASE tool and this encouraged us to collect more empirical data that would shed more light on the authority's 'clients', including actual case data and statistics. This enabled us to evolve our assumption personas even further while grounding aspects such as user activities, aptitudes, attitudes, motivations and skills in empirical evidence. Personas were elaborated even further during the persona calibration exercise. This task ran in parallel with the main project requirements workflow.

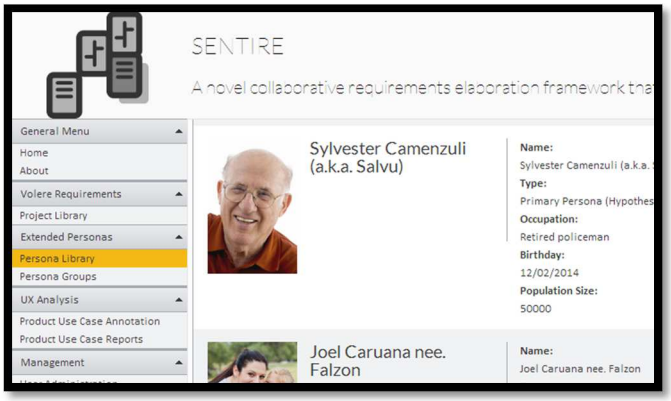

Fig. 3. Persona library in Sentire's CASE tool

Initially four primary and four secondary personas were specified:

- *Mary Piscopo* – 55-65 year old tech-newbie
- ***Joel Caruana*** – 30-40 year old teacher and mother
- ***Noel Caruana*** – 30-40 year old engineer
- ***Shanya Borg*** – 16-18 year old student

Secondary persona (won't be directly affected)

- *Joe Grech* – 55-65 year old trader (not a target user)
- *Sylvester Camenzuli* – 70+ year old retired policeman
- ***Joanne Bonnici*** – 18-25 year old content manager
- ***Joseph Zammit Borda*** – 30-40 year old case-officer

A number of user models have been previously generated in other case studies (for similar personas to those shown in bold) and thus can be reused. No models were available for *Mary Piscopo.* We organized a number of user group calibration (UGC) sessions with participants falling under this user archetype. Seven individuals accepted to participate. All ethical considerations recommended by UCL's Research Ethics Committee were taken. Calibration sessions took around 50 minutes to complete.

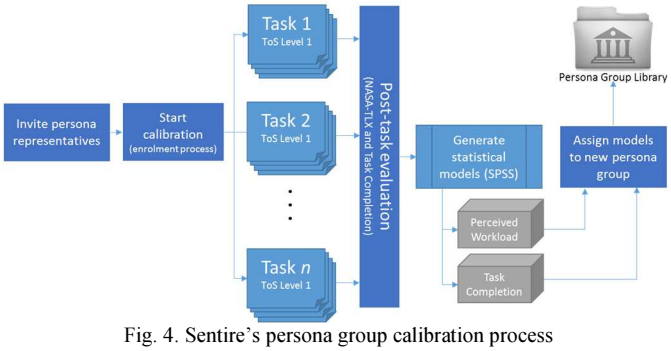

Fig. 4. Sentire's persona group calibration process

The calibration process requires participants to go through a set of nine fictitious enrolment tasks covering a wide spectrum of real-world e-service enrolment configurations, ranging from

easy (e.g. requiring users to fill in two fields and create a password) to hard (e.g. requiring users to travel physically to a government office to verify their identity as well as creating multiple credentials). After each task participants had to rate six sources of workload (*mental*, *physical* and *temporal demand*, *own performance, frustration and effort*) and specify their willingness to complete the enrolment process in four different e-service scenarios (i.e. services of increasing importance/compulsion and frequency of use). The online and self-administered version of the calibration process was not practical with this group of people mainly due to a low level of confidence in using online services for the first time. We devised a new approach whereby the facilitator explained each task using visual and verbal cues while calibration feedback was recorded on their behalf. This approach might pose a risk to validity due to potential bias (as opposed to going through the calibration process independently), however we noted a large degree of openness and honesty in their feedback which re-assured us on the authenticity and quality of the data generated. Leading questions were also avoided to mitigate this risk. This technique also helped to uncover several unexpected insights on this user group's attitudes towards enrolment. The calibration data from each session was consolidated and prepared for processing using a statistical package (SPSS). Two regression models were created: a *linear regression model* to explain perceived workload and a *binary logistic regression model* to predict the users' willingness to adopt the e-service.

	Willingness to Use Coefficients	Perceived Workload Coefficients
B-Coefficient	5.866	3.888
New items	-0.78	No impact
Items to recall	No impact	2.183
Delays	-1.434	34.332
Interruption	-1.925	24.127
Type of Service 1	-2.339	No impact
Type of Service 2	-1.448	No impact
Type of Service 3	-0.718	No impact
Type of Service 4	No impact	No impact

Fig. 5. These coefficients explain a user group's attitudes towards enrolment specific design elements (e.g. an additional form field increases perceived workload by 2.2%, while a delay increases perceived workload by 34.3% while reducing the probability of adoption by 1.4%). Statistical tests are conducted on each model to assess its predictive power and validity – calling for further evaluation and possibly calibration

Throughout the process we noticed that some participants behaved in a significantly different manner, even though they were theoretically accurate representatives of our persona. This led us to believe that our initial assumption persona (*Mary Piscopo*) may have been an over-generalization of that particular user group and we considered this as a sign that new personas might be emerging.

We started to generate more insights on our participants and a new set of attitudes started to emerge clearly suggesting that we're dealing with two different user groups. This new group of people appeared to be extremely afraid of "*breaking something*" even though they are willing to adopt new technologies. Another attitude was that they prefer physical demand (i.e. going somewhere) than feeling frustrated (i.e. fear of doing something wrong or lack of understanding). This was related to the issue of confidence, specifically wanting to get the job done with high confidence in the outcome.

We decided to create a new persona and associated user models to cater for this variant on the original persona (we called this variant *Doris Piscopo*), based on these new insights.

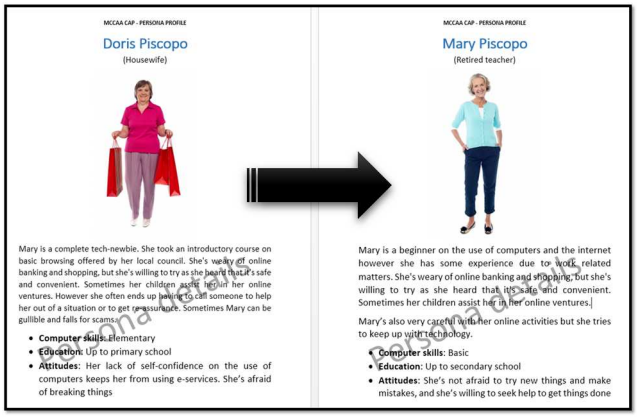

Fig. 6 Creation of new persona to reflect an emerging user archetype

A qualitative analysis of the calibration sessions confirmed our hypothesis for this new persona. The following sum this up:

- *"When I see an enrolment page I stop as I'm afraid of breaking something [the computer]"*
- *"I depend on my daughter who's still at home … when she leaves, then I'll make an effort to overcome my fear"*
- *"I prefer to go out and finish tasks in person, it's part of life and it's relaxing"*
- *"Time and physical effort are not an issue"*

"Lack of self-confidence" and "fear of breaking things" are two main attitudes associated with this new persona. These observations are grounded in feedback obtained throughout the calibration process. In the final part of the calibration process (comparison of NASA-TLX workload dimensions) it became extremely clear that this group of people are more willing to endure *physical* and *temporal demand* as long as they feel confident that they completed the job, they did it "the right way" and that they're not making any mistake that could cause harm or damage anyone or anything. One participant noted that she has a "*fear of breaking the national network [laughing]*". The calibration exercise helped us learn more about our users and

their attitudes towards specific design factors, while enabling us to operationalize and model them for future re-use.

In the meantime tasks **T3**, **T4**, **T5** as well as **P1** provided us with a solid foundation to scope the new e-service (**T7**) while evaluating possible product use cases, or rather functions that the new e-service may afford to stakeholders, primarily consumers (**T8**). Following Volere's templates, each product use case was discussed and scenarios (normal and alternative) were drafted (**T9**). All of the personas were present in a visible location during these sessions.

Fig. 7. Persona posters were highly visible during meetings

Given the required agility for this project, a hi-fidelity prototype was used to guide the team in the requirements specification and design process. We believe that in projects where high agility is required, product use case and scenario development are in themselves a design activity, rather than just requirement elicitation and specification techniques. As part of the product use case specification exercise we conducted a card-sorting exercise to determine an initial take on the e-service's information architecture. A set of yellow post-it notes representing functional and informational pages were provided together with a set of blank green ones. The group was asked to use the green notes to create categories under which the other post-it notes would be placed. This would in turn translate into a consensus based hierarchical representation of information and service pages. Following a team effort and based on experience and expectations the team managed to come up with new informational and functional categories for the e-service, while existing ones were removed or consolidated. The card-sorting exercise was crucial to reflect on specific requirements: *how will a user look for advice? Will consumers follow a hierarchical navigation pattern through FAQ drill-downs?* This shaped our idea of flows and an initial design started to emerge. Participants also had to agree on naming conventions for groups of concepts, and this informed the creation of menus, menu items and layouts (e.g. "*News*" vs "*Alerts*"). We planned to use closed card sorting (with a predetermined set of categories) however the team opted to create new categories and concepts where necessary. This gave way to discussions about user-journeys, defining the path a user would take to complete a task (e.g. seek advice or file a complaint). Using a marker the team joined various post-it notes with arrows, indicating flow. At each point throughout the discussion we referred back to our personas to determine whether we are excluding any specific user groups. The guiding principle was to design for the lowest-denominator in terms of skills. Whenever a design decision was taken we referred back to our personas and asked questions such as: "*Would it be intuitive to Joe? Would it be too complex to Mary? How would Shanya react to this?*" This was however based on subjective interpretation. Following Faily and Fléchais' [9] recommendations we attempted to back our assumption personas with facts about real consumers who have interacted with the authority. Grounding was also based on first-hand experience by stakeholders.

Fig. 8. Left: card sorting exercise following an initial iteration of product use case designs. Right: second stage of the card-sorting exercise was to determine user journeys

User journeys convey important information to the design team as they uncover flaws in assumed workflows and allow for an early rethink of the steps required to complete a primary task. This exercise should be conducted with both users' goals as well as the authority's goals in mind. At this point we had a set of product use cases, built following a systematic investigation of the authority's work and a detailed scoping exercise. Certain requirements started to emerge, and a note was taken (although no formal specifications were made at this point). We proceeded to test the current product use cases to determine their impact on users (**T10**, **T11**). Various advantages and disadvantages of introducing enrolment were discussed, both from a users' point of view but also from MCCAA's point of view. Two enrolment approaches were considered; either via the national e-ID infrastructure or via a custom-built and internal user management facility. These scenarios placed different demands on users, and the impact from each scenario was assessed individually. From a users' perspective we considered physical workload (i.e. national e-ID requires users to visit a registration office in person) and cognitive workload (i.e. custom MCCAA user accounts would require users to create yet another set of credentials). Sentire was adopted to simulate and visualize the impact that enrolment can have on users. Normal case scenarios were created for both enrolment options and these were then annotated (**T10**) as shown in Figure 9. Each use case is specified using scenarios, which are in turn specified as consecutive steps. Enrolment-specific steps are annotated with measurements for each of the design factors used in calibration. This will allow us to generate predictions using our statistical user models.

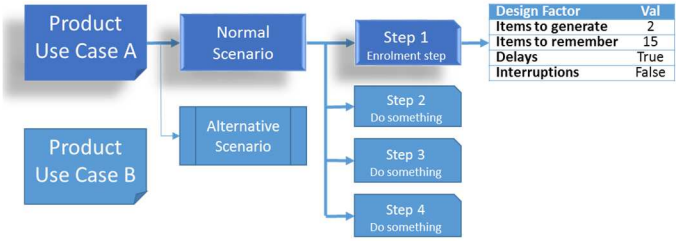

Fig. 9. Sentire's product use case annotation workflow

We then moved on to simulate user feedback through Sentire for the several user groups involved and for the different enrolment scenarios (**T11**). Before computing simulations the available user models were associated to the project's primary personas as shown in Figure 10.

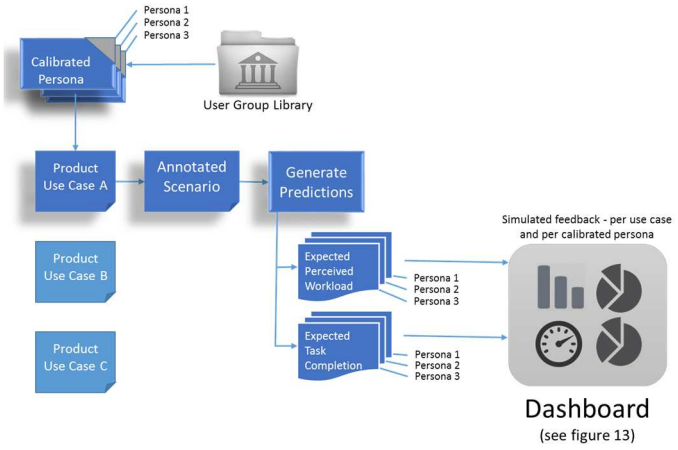

Fig. 10. Sentire's user feedback simulation workflow

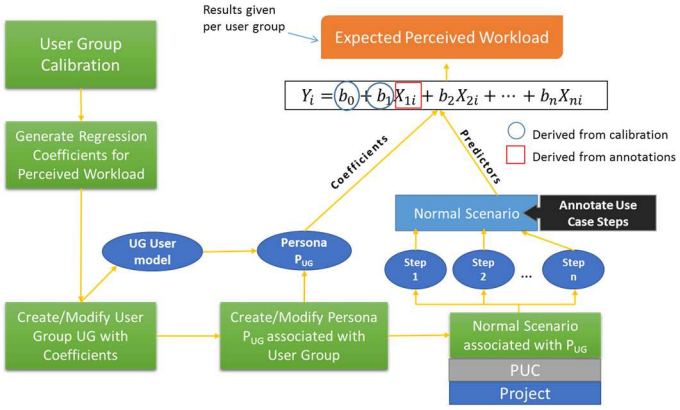

Fig. 11. Sentire's simulation algorithm for perceived workload (willingness to complete the task is calculated using a binary logistic regression function as shown in Figure 12)

$$P(WtCT) = \frac{1}{1 + e^{-(b_0+b_1X_1+b_1X_1+\cdots+b_nX_n)}}$$

Number of users (per persona) who would willingly complete the task
= P(Success) x Persona population

Fig. 12. Willingness to complete the task (WtCT) uses a binary logistic regression function

Personas are associated with statistical user models generated through calibration. These models denote regression coefficients that explain the users' reaction towards specific design elements (e.g. an additional form field will add 10% to the perceived workload and reduces the chance of task completion by 3%). These values, together with the product use case measurements (annotations) are parameterized into their respective regression functions. This provides us with predictions for perceived workload levels and willingness to complete the primary task (as opposed to giving up at enrolment stage). The simulated feedback confirmed the team's concerns and also strengthened their conviction that adopting an enrolment process based on the national e-ID for basic and infrequent transactions is an overkill, which would then result in major adoption issues. We were operating on the assumption that the number of active e-ID accounts in Malta is relatively low (i.e. around 10%). Using an MCCAA specific account would result in generally improved process completion rates (sign-ups) however this has a negative impact on users represented by *Mary* and *Doris Piscopo* while discouraging many younger users represented by *Shanya Borg* (16-18 year old users). Sentire confirmed the team's gut feelings and decisions could be taken with higher confidence, backed up with objective data.

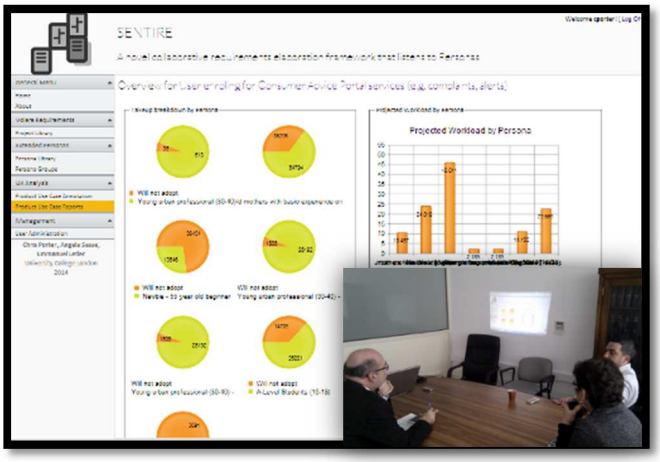

Fig. 13. Simulated feedback for the different user groups represented by the various calibrated personas used throughout the design process. The feedback shown above was generated using our Sentire CASE tool on the "*enrol with an MCCAA account*" product use case (Inset: Sentire workshop participants)

For each group of users Sentire reported (see Figure 13) the amount of perceived workload (*histogram*) as well as the willingness to enrol (*pie-chart*) for the e-service and complete the task online (as opposed to using conventional methods to file complaints and seek advice via phone, email, snail-mail or in person – or give up on the process altogether). Following a couple of iterations (changing enrolment process parameters) it was decided to leave the e-service open across all use cases without the need for any compulsory enrolment. Nonetheless an optional MCCAA account was considered to be a reasonable offering for those consumers who would wish to track their complaints and other interactions.

Persona calibration lets us capture, model and store user behaviour which can then be re-used via Sentire's persona library. Although not necessarily precise, simulated feedback challenges designers to re-assess their assumptions and decisions following a systematic and repeatable process. At this point a set of product use cases were formulated and tested for critical enrolment-related issues. Atomic requirements were specified for these use cases adopting a modified version of Volere's requirements Snow Card template (**T12**).

These low-level requirements covered various categories including *functional*, *usability and humanity*, *look and feel* and *maintainability and support*, all of which are testable, measurable and traceable. Each requirement was assigned to specific *Product Use Case/s* or marked as *Global.* This offers various levels of requirements granularity, with the product use

case being a high level view of what the system shall do, and atomic requirements specifying low level detail denoting how a system shall achieve such functionality.

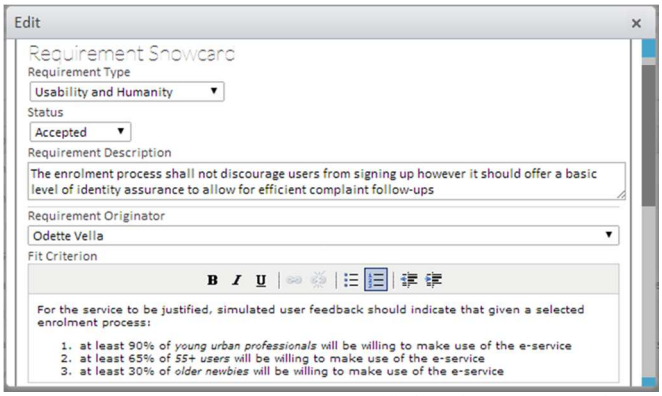

Fig. 14. Atomic requirements were specified for these use cases adopting a modified version of Volere's requirements Snow Card template (and using measurable UX-related fit-criteria).

Sentire's CASE tool offers a project visualization map, allowing the team to view requirements at various levels of granularity as well as their inter-relationship with higher level groupings (e.g. product use cases, business use cases and events). Other project elements are also displayed (and linked), including stakeholders, personas/actors, events, risks, use cases and requirements. This can serve as a visual impact assessment utility for regression testing following modifications to requirements or use cases.

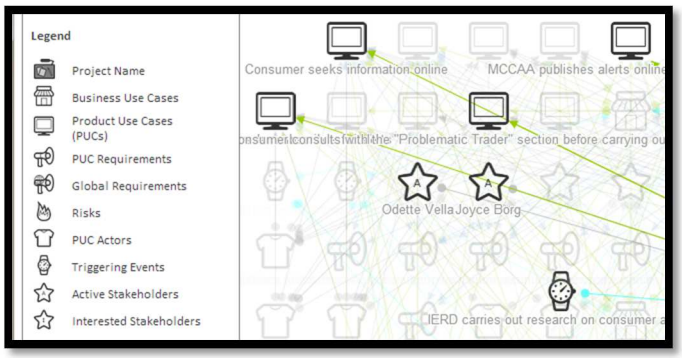

Fig. 15. Sentire's CASE tool: project map – indicating relationships between different elements within a project.

Based on the current level of coverage and detail, a prototype started to emerge (**T13**). The hi-fidelity prototype was tested with a number of participants at University of Malta's Usability Lab (**T14**). Using Tobii's eye-tracking and analysis studio, a set of pre-determined tasks (goals) are provided during which eye-gaze data is captured for a deeper assessment on findability, navigability and explicit pain-points (e.g. using heat-maps to uncover points of failure). Retrospective Think Aloud (RTA) sessions provide us with deeper and invaluable insights and knowledge on what users expect, what they look for and the rationale behind their decisions. This data supplements the eye-tracking information.

A number of severe usability issues were uncovered during the first few sessions. These were corrected prior to the subsequent sessions. Following this iterative process the authority can now plan case officers' training, data migration and release.

Fig. 16. Prototyping the e-service based on the initial information architecture session

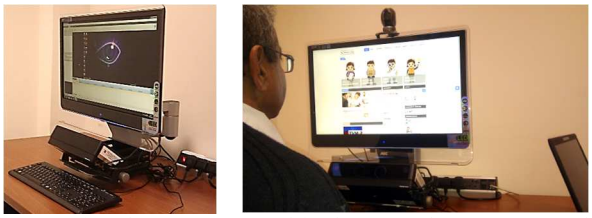

Fig. 17. Left: Tobii X-120 was used for the eye-tracking sessions. Right: Participant during an eye-tracking session

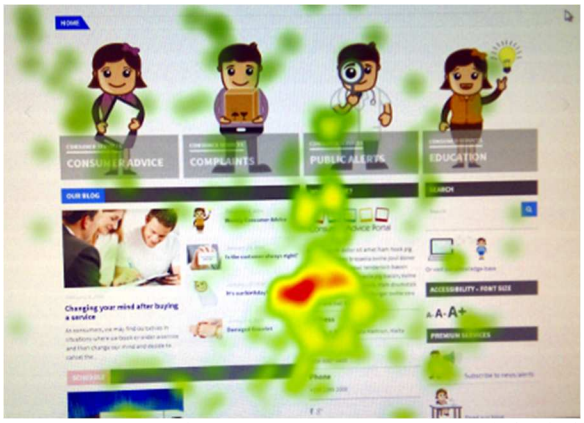

Fig. 18. Heat-maps indicate common gaze patterns across individuals

IV. Conclusions and Recommendations

A. Persona Evolution through Calibration

An interesting side-effect of the persona calibration exercise was the identification of new personas stemming from behavioural information. Our primary persona *Mary Piscopo* was built on assumptions, primary data, observations, opinions and experience provided by the stakeholders. When representatives of this persona were calibrated it was immediately clear that there were several distinctive behavioural aspects between various participants who were theoretically similar in demographic terms. This informed the evolution of *Mary Piscopo* but more importantly to the creation of a new persona *Doris Piscopo* which reflected the contrasting attitudes towards specific enrolment-related design factors. Although a similar result could have emerged through ethnographic methods we believe that the systematic, structured and objective nature of the calibration process highlighted marked differences

between expected and actual behaviour across participants. When distinct clusters of behavioural information emerge from the data then this would be indicative of a phenomenon that requires further investigation (e.g. out of seven participants, three generated commonly divergent results with respect to perceived workload and willingness to complete a task). A small number of participants may not yield enough information to generate statistically significant results (models), however these are all important insights that could drive the team to consider different angles of the same problem while being critical of their own actions and conclusions - promoting a structured and actively reflective design process. An unstructured and qualitative discussion with participants might not have the same effect as the systematic and repeatable user group calibration (UGC) process.

B. Partial User Models

Sometimes user models can only partially capture user behaviour and reactions towards specific design factors. This is mainly because certain predictors used during the calibration process would not be statistically significant for a specific group of users (e.g. for a given user group and in a given context, an increasing amount of fields to fill may have an impact on workload but not on the willingness to complete the task). Another major cause of partial user models is when data collection is not extensive enough to perform good model fitting (using linear or logistic regression) and the influence exerted by certain predictors would degrade the model's overall predictive power. Statistical tests would highlight these predictors which are in turn excluded from the final model to a) strengthen the effect of the remaining predictors and b) improve overall predictive power, albeit for specific design factors only. This however must be made explicitly clear to avoid misinterpretation. Figure 19 shows a conceptual representation of how this can be communicated.

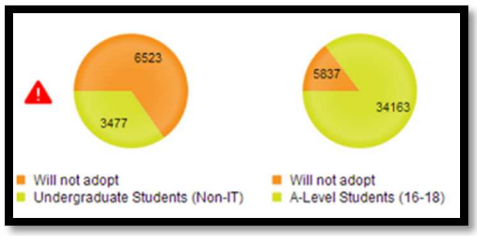

Fig. 19. A conceptual representation of alerts shown when simulations are generated via partial user models (i.e. first chart is based on a partial model).

C. Indicative can be as Good as Precise – as long as it's comparable

The team described the experience as intuitive and non-threatening to the non-technical person. This is mainly due to its technology-agnostic Volere-based process. Simulated feedback provides a truly user-centric and objective grounding for discussion and from the experience gained throughout this particular project it was clear that the team was not interested in precise predictions but considered general trends or indications to be sufficiently useful to inform decision making. Management felt more comfortable taking decisions based on a quantitative (and comparable) view of the impact that certain design decisions would have on users. The introduction of a new calibrated persona (*Doris Piscopo*) was a case in point where its behavioural simulations were only indicative and based on a weak statistical user model. This was mainly because data for the underlying user group (*complete newbies (55+)*) was based on readings from only three UGC participants. Nonetheless this calibrated persona was still useful as it gave an objective, yet broad indication of what kind of reaction to expect in comparison to the other calibrated personas. Additional calibration sessions would strengthen the user model underlying this persona. Nonetheless the current, albeit weak model still contributed towards a design decision, that of eliminating all enrolment processes from the e-service.

D. Contextual Feedback is Possible with In-Context Calibration

Persona group calibration can be carried out either in a lab environment or within the users' natural environment from where the e-service will be used. In an earlier case study we calibrated participants representing the *young urban professional (30-40)* user group. In this case calibration was carried out at the participants' workplace and it was noted that decisions were highly influenced by contextual nuances. Participants could refer to their physical surroundings and work-conditions before submitting their feedback on perceived workload and task completion (e.g. "*how hard would it be to scan a utility bill over here?*"). Behaviours and attitudes might change in different contexts, however in-context calibration may mitigate this risk by creating behavioural models influenced by contextual nuances.

E. Re-use

Knowledge gained throughout the process, including terminology, stakeholders, personas, behavioural models and requirements have been stored in the online CASE tool and may be re-used for future projects by MCCAA or possibly by other national and/or regional entities. This reduces the *cold-start problem* in future projects since accumulated knowledge would be available for immediate re-use – potentially resulting in less time spent on aspects such as user group calibration. Some critics have argued that the expense to calibrate personas could be avoided and resources should be used to test e-services directly with users. We believe that Sentire is not intended to replace user testing but its purpose is to assist design teams from the earliest stages of an e-services' life-cycle (through simulated UX-analytics) to mitigate against bad design decisions on critical success factors (e.g. enrolment). A systematic method is provided to build a grounded understanding on different user groups. This contributes towards a persona-based decision support system integrated within an industry strength requirements process. We strongly believe that Sentire's value increases as more user-specific knowledge is accumulated and maintained for use within future e-services.

V. Future Work - Toolset

Stakeholders at MCCAA were too busy to actively engage using the online CASE tool. This required email-based updates and calls for feedback. All correspondence was stored in the

tool's project repository (as HTML documents). We believe that *multi-channel alerts* would encourage engagement. Stakeholders could be given the option to select their preferred way/s to receive alerts in case any element within the project is modified (e.g. SMS notifications when specific requirements are modified and email notifications for changes in the terminology section).

Meeting notes were taken in an online notepad made available in the project workspace, however following a number of meetings this got too large and impractical to use for reference purposes. A project *meeting management area* could be introduced – offering meeting recording facilities (meeting notes and meta-information) as well as a meeting-specific repository for artefacts collected during each meeting (e.g. photos, recordings, and documents).

Use case annotation and feedback simulation should be made available on a *sidebar* within the use case formulation page. At the moment user feedback simulations are only available from a separate area within the tool and thus creating a break in the workflow. A realtime *chat* facility would offer a centralized space for teamwork. Some stakeholders prefer using email, and an additional feature could allow users to send *emails directly* into Sentire's CASE tool by providing project-specific addresses, such as mccaa_project@devbell.com. Wireframes are currently created using external tools and images are then imported into the CASE tool. We're working with a wire-framing app vendor to create an *integration API to offer in-app collaborative wireframing* capabilities.

More work is also required to study the effectiveness of and to improve the calibration process for use with non-technical participants.

ACKNOWLEDGMENT

We would like to show our gratitude to the Malta Competition and Consumer Affairs Authority (MCCAA) for their time and support. We would also like to thank all the persona calibration participants for their time and patience.

REFERENCES

[1] Jokela, T., & Buie, E, "Getting UX into the contract," in E. Buie, & D. Murray (Eds.), Usability in government systems: User experience design for citizens and public servants (pp. 251-266) MK – Elsevier, 2012.

[2] Lehtonen, T., Kumpulainen, J., Jokela, T., & Liukkonen, T. Nm "To what extent usability truly matters? A study on usability requirementsin call-for-tenders of software systems, issued by public authorities" in Proceedings of NordiCHI 2010: Extending Boundaries, Reykjavik, Iceland. pp. 719-722, 2010.

[3] Janowski, T., Estevez, E., & Ojo, A, "A project framework for e-government" in Proceedings of the 4th International Conference on E-Government, Copenhagen, 2005.

[4] Robertson, S., & Robertson, J. Mastering the requirements process: Getting requirements right. Addison-Wesley, 2013

[5] Nielsen, Jakob. "Usability Metrics". Nielsen Norman Group. http://www.nngroup.com/articles/usability-metrics/, 2001 (accessed March 2014)

[6] S. Hart G. and L. Staveland E., "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," Advances in Psychology - Human Mental Workload, vol. 52, pp. 139-183, 1988.

[7] C. Porter, A. Sasse M. and E. Letier, "Designing acceptable user registration processes for e-services," in Proceedings of HCI 2012 - The 26th BCS Conference on Human Computer Interaction, 2012.

[8] A. Cooper, R. Reimann and D. Cronin, About Face 3: The Essentials of Interaction Design, Wiley, 2007.

[9] S. Faily and I. Fléchais, "Persona Cases : A Technique for Grounding Personas," Proceedings of the 29th International Conference on Human Factors in Computing Systems, pp. 2267-2270, 2011.

[10] C. Porter, M.A. Sasse and E. Letier, "Giving a voice to personas in the design of e-government identity processes," in Research to Design: Challenges of Qualitative Data Representation and Interpretation in HCI - in BCS HCI 2013, 2013.

[11] Nielsen, Jakob. "Usability Testing of WWW Designs: Article by Jakob Nielsen." Nielsen Norman Group. http://www.nngroup.com/articles/usability-testing-1995-sun-microsystems-website/, 2001 (accessed February 25, 2014).

[12] Trigg, Randy, and Andrew Clement. "Participatory Design." Computer Professionals for Social Responsibility. http://cpsr.org/prevsite/program/workplace/PD.html/, 2000 (accessed March 6, 2014).

[13] Porter, C, "SENTIRE / designing human-centric identity systems". Poster session presented at the Information Assurance Advisory Council (IAAC) Symposium, London, 2013.

[14] Faily, S, "A framework for usable and secure system design", University of Oxford, 2011.

[15] Alexander, Ian, and Suzanne Robertson. "Understanding Project Sociology by Modeling Stakeholders." Stakeholders without tears. http://www.scenarioplus.org.uk/papers/stakeholders_without_tears/stakeholders_without_tears.htm (accessed March 6, 2014).

[16] K. Pernice and J. Nielsen, "How to Conduct Eyetracking Studies," Nielsen Norman Group. 2009.

[17] HM Revenue & Customs, "Online services: digital certificates", retrieved May 2012, from HM Revenue & Customs: Online Services: Digital Certificates: http://www.hmrc.gov.uk/ebu/digital-certs.htm, 2012.

[18] OECD, "Security and Authentication Issues in the Delivery of Electronic Services to Taxpayers". Retrieved October 2012, from http://www.oecd.org/, 2012.

[19] Egovbarriers.org, "Deliverable 3: Solutions for eGovernment - Section 3: Solutions for the seven eGovernment barriers", retrieved May 2012, from Breaking Barriers to eGovernment: http://www.egovbarriers.org/, 2007

[20] Martens, T, "Electronic identity management in Estonia between market and state governance" in Identity in the Information Society, 2010.

[21] ISO. "ISO DIS 9241-210: Ergonomics of human-system interaction – Part 210: Human-centred design process for interactive systems". ISO, 2008.