Learning
discovering what users want.

Community
Preferences
connecting what users prefer.

User Intent
predicting what users will do.

Community
Orchestration
organising users’ connections.

Trust
connecting users and reputations.

Privacy
managing personal data.

Community
Context
connecting users and their environment.

Location
providing rich location data.
©Societies 2013

Disclaimer:
All SOCIETIES consortium parties have agreed to full publication of this document.

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

Neither the SOCIETIES consortium as a whole, nor a certain party of the SOCIETIES consortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

SOCIETIES is funded by the EC under FP7
A Pervasive Community is inherently context-aware, self-organising, self-improving and capable of pro-active behaviour aiming to optimise and personalise the pervasive experience of an entire community. In addition to the resources controlled by its individual members, a Pervasive Community may also provide public access to its devices, services and resources.

The notion of Cooperating Smart Spaces (CSSs) has been introduced so as to extend pervasive systems beyond the individual to dynamic communities of users. The SOCIETIES functionality can be categorised in terms of three broad phases each of which contributes to the formation of CSSs which are: Discover, Connect and Organise.

SOCIETIES enables the Discovery, Connection and Organisation of relevant people, resources and things, crossing the boundary between the real and virtual worlds.

**SOCIETIES Objectives**

The goal of SOCIETIES will be achieved through four key objectives:

- To facilitate the creation, organisation, management and communication of communities via Cooperating Smart Spaces, where pervasive computing is integrated with social computing communities;

- To provide an enhanced user experience for both individuals and entire user communities, based on proactive smart space behaviour and dynamic sharing of community resources across geographic boundaries;

- To design and prototype a robust open and scalable system for self-orchestrating CSSs;

- To evaluate, through strong involvement of end-users, the usefulness and acceptance of the developed CSS software via three user trials with the three distinctive groups.

**SOCIETIES user groups:**

- Enterprise Users: Enterprise communities play an important role in bringing together people, goods and services within global markets, local ecosystems or large organisations. The CSS concept will bridge the gap between smart IT systems and established enterprise community activities.

- Students: Students adapt easily to new technology, and since communication and social networking play an integral role in their lives, they are most likely to adopt CSSs, using them in ways both foreseen and unforeseen.

- Relief Experts: The ability to rapidly form a disaster management community from all the closely located relief teams can help save lives, property, and the environment.
Content

Community Preferences
10

Trust
16

6 Innovation Areas
8 Learning
10 Community Preferences
12 User Intent
14 Community Orchestration
16 Trust
18 Privacy
20 Community Context
22 Location
24 Contact Information
Discover, Connect, Organise
Innovation areas

To deliver on such a bold SOCIETIES Pervasive community vision many diverse areas must be addressed for example, privacy and learning. The following articles outline some of the more innovative aspects of the system and an application of the innovation is also highlighted. A broad overview of the innovations covered in each article is given below.

Learning

Learning enables SOCIETIES to acquire knowledge about individual end-users and communities of users by monitoring their actions over time. Whilst in most social media, such knowledge would be used to generate revenue via targeted advertisements, etc. in SOCIETIES the driver is to help the end-user invoke and personalise relevant system components and third party services. Learning obviates the need for end-users to enter all this information manually and the community aspect enables each end-user to capitalise on the knowledge captured about other, similar, end-users. Learning underpins many of the other innovations in SOCIETIES.

Community Preferences

Community preferences enable SOCIETIES to build templates, or stereotypical sets of preferences for a whole community based on the preferences of individual community members. This can be particularly helpful to new members of a community who may “inherit” all, or a subset of, the preferences of a community when they join and benefit from the “wisdom of the crowd”. SOCIETIES is also able to derive community preferences from hierarchies of communities. Current social media do not support community hierarchies or the inheritance of properties from the circles/groups/networks etc. to which an end-user might belong.

User Intent

For any system to react appropriately for a given end-user, it needs to be aware of what that end-user is attempting to achieve i.e. the ultimate goal of the end-users actions. In systems which allow end-users to interact freely and navigate their own way through the available options, user intent is not easy to deduce and current social media make no attempt to do so. SOCIETIES captures this knowledge by monitoring end-user behaviours and the context in which those behaviours occur. Observing temporal sequences of end-user actions and context cliques or snapshots can permit the discovery of past, and prediction of future, goals.
Community Orchestration

The richness of the pervasive communities supported by SOCIETIES presents many new challenges for the end-user in discovering, managing and organising the communities to which they belong. This is primarily due to the physical resources (devices, sensors, services) of pervasive communities. These add "real world" impact to a community in SOCIETIES which current social media are only just beginning to explore. In order to maximise the value to end-users, community orchestration helps end-users to identify and discover relevant communities and manage the intelligent formation, organisation, membership and termination of communities.

Community Context

Context awareness lies at the heart of pervasive computing in general and the pervasive communities of SOCIETIES in particular. Current context models do not support the management of context for dynamic communities of individuals in large scale systems. Community context in SOCIETIES supports context conflict resolution and context inheritance for communities thus further enabling end-users to benefit from the "wisdom of the crowd". Furthermore, SOCIETIES is not only able to form communities based on end-users with similar context, it can also contribute to the life-cycle management of these communities as changes occur in end-user context.

Trust

As the number of contacts, pervasive services and devices available to a given individual proliferates it will become ever more important for end-users to assess the trustworthiness of the entities with whom they interact. SOCIETIES supports robust and authenticated trust assessment mechanisms. Aspects of trust (e.g. purposeful or referral) are formalised in SOCIETIES and extend well beyond the recommendation systems of current social media.

Privacy

SOCIETIES has been designed and developed with a "privacy by design" approach. This means that privacy protection is fully integrated into the platform rather than being an add-on. The privacy protection systems provided by SOCIETIES help the end-user to manage their personal data and its disclosure. SOCIETIES supports, and enforces; negotiation between end-user preferences and service/community policies; data disclosure with personalised obfuscation; multiple identity management and selection; and privacy assessment and auditing.

Location

Of particular importance amongst the rich set of context attributes handled by SOCIETIES is location. Social media have begun to take advantage of this by enabling end-users to inform others of their location automatically and to "check in" to popular locations. The full benefits of location based services, however, have yet to be realised because of the lack of precision in the data, particularly for indoor locations. SOCIETIES supports a variety of locating systems, of which "WiFi sniffers" are perhaps the most innovative. These can pinpoint indoor locations very accurately by independently monitoring the signal strengths between end-users and access points.

ASIDE:

Concepts: A supporting concept for making this possible is called Community Interaction Space (CIS). It enables a group of individuals, their devices, services and resources to be dynamically bound together based on particular context attributes – such as location or common interests. The CIS allows individual community members to share resources, personal information, context data, services and devices, in a privacy aware manner, with other members of that community.

The notion of Cooperating Smart Spaces (CSSs) has been introduced to represent an end-user, or other stakeholder. It allows them to manage their multiple sensors, devices, services, resources, and information in a convenient integrated way, as part of a single CSS. The real benefit can be seen when these individual entities form new or join existing communities to collaborate.
Personalisation

Learning

Learning enables SOCIETIES to acquire knowledge about individual end-users and communities of end-users.

ASIDE:
A Pervasive Computing system can be defined as “machines that fit the human environment instead of forcing humans to enter theirs”. It is the hardware and software required in order to fulfil the “invisible computer” paradigm; where information processing has been thoroughly integrated into everyday objects and activities; and is end-user focused. It is driven by a high-level conceptual model consisting of devices, end-users, software components and user interfaces. This allows the system to be context-aware ie. aware of physical, virtual and informational situation in which the system is being used.

Innovation area
Learning in social media tends to be for the benefit of the service provider, such as data mining to provide revenue streams (e.g. personalised adverts). This only scratches the surface of the possible benefits, and provides limited benefits to the average end-user of a service.

Learning in pervasive computing systems is directed more to supporting the end-user by implicitly acquiring information which they would otherwise have to supply explicitly. Typically only individual end-users are considered. Learning at a community level is completely new.

Innovation: Implicit personalisation
Learning in SOCIETIES can assist the end-user in selecting and configuring the multitude of services and devices with which they can interact in a pervasive world. To undertake all this manually in an explicit manner would require significant time and effort. Learning enables SOCIETIES to acquire all this information implicitly by monitoring the end-user’s choices and behaviours and learning what they prefer in a context-aware fashion.

In more technical terms, SOCIETIES employs two main learning mechanisms; one with a long time constant and one with a short time constant –

1. A batch learning algorithm (C4.5) which is run on a periodic basis to data mine a history of the end-user’s activities;
2. A dynamic incremental algorithm (DIANNE) which reacts directly to an end-user’s actions and so learns new choices/behaviours immediately.

Both of these learning systems can be applied to communities of users as well as individual users (see also Community Enhanced Personalisation).

In situations where an end-user (or community of users) is likely to prefer a more considered approach to learning, i.e. where rapid learning is not required so a more detailed analysis of their behaviour over time can be undertaken then batch learning can be employed. SOCIETIES uses traditional supervised learning methods such as C4.5 for this. However, there are also times when an end-user would like the system to react more quickly to a change in their behaviour and this is where incremental learning comes into its own. SOCIETIES uses a novel neural network approach to achieve this.

Applied Scenario
Every time Bob travels to meet a client, he carries out the following steps. He books a flight online, a hotel with “hotels.com” and a return train ticket to
A Pervasive Computing system can be defined as "machines that fit the human environment instead of forcing humans to enter theirs".

the airport. The day prior to departure he synchronises his laptop with the document repository so that the latest marketing material is available to him, and prints out a copy for the client. After several trips the system recognises the "sales trip" sequence using the C4.5 algorithm, and can partially automate the process when the sequence of steps is unambiguous (e.g. after the hotel is booked).

Last week Bob has become aware of reducing print costs at work, so he now emails a PDF of the marketing brochure to the client instead of printing one. This requires the incremental learning algorithm to detect that the trained behaviour of automatically sending the document to the printer, is now incorrect.

Innovation: Stress recognition

The SOCIETIES system can learn to detect stress in individuals. A probabilistic mathematical model is used to capture the relation between end-users' bio-signs and stress levels (see diagram below). Inputs from bio-sensors (for example, blood pressure and pulse monitor) attached to the individual end-user are processed in real-time to provide an inferred stress level which is stored as contextual information about the end-user.

By applying a "Stroop" Colour Test five different stress levels are simulated. To train the system this test is used in a learning environment by creating artificial stress situations for which the biophysical reactions can be measured. The test subjects are asked to gauge their stress levels and this self-assessment is used as "Ground Truth" for stress levels. Automatic learning algorithms can then produce the mathematical model (Bayesian Network) used to automatically determine the "stress" of an individual.

Applied Scenario

Stress level information is used by other intelligent systems within the SOCIETIES platform for decision making. For example, stress recognition is particularly relevant for end-users in dangerous situations. It ensures that they do not get disturbed by irrelevant messages during stressful situations, e.g. a driver in dense traffic. In disaster management it can be used for relief work assignments, as many helpers work far beyond their expected capacity. Often they do not even notice their higher stress levels, due to the high motivation to help and the importance of the work. However, if their stress level was not taken into account, their error rate would increase significantly and they could potentially put themselves and others at risk.
Community Preferences

Combining the preferences of the members of a community.

ASIDE:
A “User Preference” can be thought of as an individual’s personal preferred option from a set of possible alternatives. These preferences can then be taken into account in automatic decision making processes. For example an end-user, when given a choice of fruit between apples, oranges and bananas, may select “oranges”. Thus, the end-user is said to have a “fruit-choice” preference for “oranges”. However, preferences can change depending on the user’s context. For example, whilst the end-user might prefer oranges in the summer, they may prefer apples in the winter. A User Preference can accommodate this “context dependency” by formulating preferences as rules. For example, IF season=summer THEN fruit-choice=“oranges” ELSEIF season=winter THEN fruit-choice=“apples”. User Preferences are not limited to propositional assertions; they can also record preferred habits or actions. For example, IF time-of-day=night THEN turn-on-the-lights. In a complex system such as SOCIETIES, manually entering all of one’s preferences could become prohibitively arduous and time consuming. SOCIETIES mitigates this by monitoring the choices and behaviours of a user and storing them, along with the context in which they were selected, for future re-use by employing standard machine learning techniques and our own specially devised DIANNE dynamic incremental artificial neural network.

Innovation area.
Community Preferences provide a mechanism for aggregating the individual preferences of all members of a community. The combined set of common preferences is associated with the community and each individual member can then inherit all, or some, of that preference set. This greatly increases the convenience of using the SOCIETIES system, as the alternative is to require all end-users to manually enter every preference for themselves. In a large scale deployment, without this functionality, it is likely that manually entering all of the preferences required to make the system useful, would place too great a burden on the end-user and, somewhat ironically, they might prefer not to take advantage of the personalisation capabilities of SOCIETIES at all.

Innovation:
Community enhanced personalisation
Historic choices, behaviours and context data are collected from all members of a community (according to member privacy settings) and fused together to create a single history for the entire community. The community history is processed offline by batch machine learning techniques to extract context-dependent “Community Preferences” which become associated with the related community.

Individual community members can inherit all, or part, of the community preference set to enhance their own preference set. Furthermore, new community members can inherit all, or part, of the community preference set when then join a community to provide an initial default preference set which they can subsequently refine. Community preference inheritance allows end-users to benefit from the “wisdom of the community” whilst still being able to manually personalise their experience with as many, or as few, personal changes to this set as they desire. This adds considerably to the convenience of using SOCIETIES, especially for new community members.
Applied Scenario
Simon is a new employee at the ACME company. During his induction he is given a SOCIETIES enabled device and is signed up to two communities; “ACME Employees” and “Customer Service Department”. In the Customer Service Department there are meeting rooms with wall displays that can be customised to a given client’s name and logo when a meeting occurs. Simon accepts the suggested default preferences from his departmental colleagues in the “Customer Service Department” community, which set an appropriate background colour and logo size for the display of a client’s logo. He also sees a preference for arranging client meetings in the morning, in keeping with the ACME company’s afternoon “quiet time” ethos and the company’s preferred finger food and refreshment option. He schedules a meeting and books the meeting room with these useful preferences which he has inherited from the community.
User Intent

For any system to react appropriately it needs to be aware of, or to predict, the ultimate goal of an end-user’s actions.

Innovation area

Most user intent learning approaches aim to predict future end-user intentions based on raw context data. This is very difficult due to the gap between the low semantic level of raw context and the high semantic level of user intentions.

User intent is not formally considered in social media. User goals have been considered in pervasive systems but predicting future behaviours has been limited to the level of the individual. No attempt has been made to apply these techniques to communities of end-users, nor to exploit/investigate the "peer pressure" effect of communities on the behaviour of individual members. This is an innovative approach that takes into account social aspects as well as more personal goals.

Innovation: User Intent prediction

We have developed three different algorithms for user intent prediction:

- Conditional Random Fields (CRFs) based user intent aims to discover and predict user intent by leveraging both the Action a user performed and the Situation a user finds themselves in (for example checking flight departure times in an airport lounge). The general process is: segmenting a Raw Context Trace of many context attributes over time into a sequence of Context Cliques; abstracting the Situations from the Context Cliques whilst monitoring the Actions of the user; and finally, determining the current intent of the user based on their current real-time Action and Situation.

- Context Aware User Intent (CAUI) provides the necessary mechanisms for modelling and predicting actions performed
by an end-user by grouping frequently occurring sequences of User Actions into User Tasks. An end-user behaviour model is developed as a stochastic model that describes the sequences of User Tasks and User Actions and the respective transition probabilities. Each User Action is accompanied by a context data snapshot describing the situation of the user.

Context Aware Community Intent (CACI) aims to create a community-wide user intent model in order to assist the prediction process for individual end-users (see ASIDE). There are two major approaches for learning a community intent model. These are differentiated by what the system collects from each community member: a history log containing end-user’s actions (see case A in diagram); or a pre-learned intent model for the individual (see case B in diagram). In both approaches the collected data are processed by a variant of the community intent discovery algorithm which results in a community intent model.

During the prediction process, if the user intent model (calculated from the CAUI and CRF algorithms) is not providing adequate results, the community intent model is also used.

Applied Scenario
Alice is a student at “NTUA” in Athens. She studies architecture and has a report to write on the influence of classical Greek architecture. The SOCIETIES system has detected her typical work flow (a User Task) - Finding a quite spot in the library, opening up her online curriculum portal “Moodle”, finding the assignment brief, starting a word processor, and starting a browser connected to a service providing pictures of architectural features.

However, normally Alice would be having fun with her friends from the campus on Monday evenings. Two of her friends pop into the library, and try to persuade her to have some fun, after all the rest of her classmates are socialising, not writing this assignment. She gives in.

The history log of Alice’s interactions along with context information are collected in a common context history repository. This captures her typical “assignment writing” User Task. The individual user intent models from her “classmate” community are processed by the learning algorithms, to create a community-wide behaviour model. This model captures the behaviour of her classmates and friends, and offers an alternate means to enhance the prediction of Alice’s actual behaviour.

ASIDE:
“Community versus the individual”. By knowing the typical behaviour of a community member, new member’s can benefit from the existing community’s experience. The behavioural data does not exist for this new end-user, so the community behaviour model can be used as an intermediate measure, while information for a more accurate user behaviour model is being gathered. The effect which a community has on the behaviour of an individual is often referred to as “peer-pressure”, and this can be very useful in predicting the behaviour of an individual end-user.


Community Orchestration

Helps end-users to manage the intelligent identification, formation, organisation, membership and termination of dynamic virtual communities.

Innovation area
Community orchestration is the ability to help end-users manage dynamic virtual communities. This can be identification and formation of new communities, for example, identifying a subset of end-users that regularly interact and have a similar interest profile, and suggest that they form a sub-community. It also assists with the organisation of community members, for example, suggesting an end-user join a community based on matching the typical community member profile, and having a similar context situation i.e. location and time.

By inferring the "community nature" i.e. temporary or ongoing community, it is also possible to suggest when to terminate a community. For example, there are no current members, and the members who were in the community had light social ties or interactions. The "intelligence" is provided by ensuring that a certain level of confidence is associated with the suggestion. For example, if the confidence in the suggestion is greater than 90% then inform the end-user, otherwise do nothing and observe the end-user further.

Management of communities in current social media is largely manual with some assistance provided in the form of friend/associate suggestions, for example. A key innovative aspect of this area is the development of algorithms that permit the continuous evaluation of communities, individuals, and a community’s nature; so that results are available to the end-user in time to make an informed decision.

Innovation:
Context state models
Individuals and communities will generate large amounts of data continuously and this data needs to be gathered and processed so that it can be analysed in a time ordered manner.

Potential new communities, defunct communities and existing communities which an individual might wish to join/leave are identified via Context State Models (CSM). Each CSM is made up of key attributes that describe characteristics relevant to existing and potential communities.

CSMs are used for analysing in near real time group dynamics. CSM modelling allows quick computational comparisons which is vital for the discovery of potential matches between communities and end-users. This then ensures that an end-user has a 'live' (near real time) experience.

CSM's also permit end-user data to be analysed anonymously in an abstract fashion. The community orchestration system uses the platform’s privacy system to enforce the individual privacy constraints of end-users prior to any automatic or semi-automatic orchestration activity.

Applied Scenario
Andy has an interest in photography, however he is currently unsure how to get the best out of his new D-SLR camera. While he pops into a local hotel, he posts some tweets about his new camera purchase. He receives a suggestion to join the "Happy Snapper’s" photographic club.

The similarity between Andy and the "Happy Snapper’s" community is captured in a CSM. It captures the fact that Andy is at the same location, and that they share the same interest "Photography". Thus, the SOCIETIES platform has suggested that he join the "Happy Snapper’s" community. The process is event-driven and reacts to events as they occur (e.g. when Andy arrives at the hotel) to determine how group dynamics are affected by CSM.
state changes i.e. invite Andy to join as he is currently not a member of the community.

Innovation: Community Nature

The Community nature refers to the concepts of “temporary” and “ongoing” communities. Some communities form and can last for years, for example a community based on family relationships. Others are formed on demand and are discarded within minutes. For example, a community formed while waiting at a train stop allows anonymous game playing to pass the time. However, the same individuals are not always present thus the community is transient and temporary in nature.

Distinguishing between communities that are likely to be short-lived (e.g. location or purpose based) and those that are likely to be of indefinite duration (e.g. family based) can assist in efficient community orchestration and life-cycle management.

Algorithms which are specific to the ongoing/temporary nature of communities can autonomously/semi-autonomously drive the creation, configuration and deletion of communities on behalf of the end-user. The community nature i.e. the concepts of “ongoing” and “temporary”, can be exploited by the system to assist an end-user.

- Longevity of association between individuals is clearly a key driver for potential community formation and, in the online world, it may be acceptable for this to be a pre-requisite.
- However, in the physical pervasive world, we need also to support ad-hoc community formation for short-term goals which would not be possible with a longevity pre-requisite.

- Discovery algorithms can look for brief, temporary communities at short intervals and longer-lasting and ongoing ones at longer intervals.
- The data used for this and how to analyse it varies depending on how long a community might be expected to last.
- Temporal and community nature semantics can be attached to communities and used to inform decisions, such as deleting obsolete communities, configuring them and creating new ones.

This will relieve the end-user from lots of housekeeping tasks associated with the many communities to which they will belong.

Applied Scenario

Bart has a busy lifestyle. He is a South African living in Dublin for some time, and uses the SOCIETIES system to keep track of everything. He likes to check the “family” community daily to keep up to speed with his brothers and sister. His brother Bob has just got a job promotion, so he sends a congratulations message, and schedules a “Skype” call for early tomorrow morning.

He is also a member of the “Springbok” community, which keeps him up to date on what is happening around Dublin. Over the summer they meet once a week to socialise. He met Zack through this group and now counts Zack as one of his friends.

On his way to work, Bart sees the “Star-cents” community. He joins in the morning on his way to work to organise a group discount for five or more people ordering coffee in “Starbucks”. It’s never the same people, but there are one or two regulars. They take turns at coming up with fun descriptive names for the community, “Drowsy sheep society” was his favourite. Other members in the queue join at the coffee shop and the community always gets his order for “semi-skimmed Latte” just right. Coffee is great, but discounted coffee always tastes that bit sweeter.
Trust
Assisting an individual to assess the trustworthiness of the entities with whom they interact.

Innovation area
Recommendation mechanisms are widely used for reputation guidance but these are not infallible, nor are they claimed to be. More formal trust assessment requires robust and authenticated mechanisms. Different types of trust (e.g. purposive, referral) and trust relations are the subject of research but they are not used in social media.

Innovation: Community-enhanced Trust Assessment
Assessment of the trustworthiness of individuals, communities and services is based not only on the experiences of the end-user, but also on the experiences of fellow members of a Community by exploiting available feedback and using learning mechanisms. This provides the necessary facilities for trust-based discovery of individuals, communities and services, and can be exploited to provide the following features:

- Supports automatic interconnection of trusted people and resources,
- Enables the sharing of privacy-sensitive information with trusted people and services,
- Supports trust-based community membership management, sub-community formation based on the trust relationships among members of the parent community, trust-based community life-cycle transition from temporary to ongoing and facilitates the trust assessment process with community feedback.

There are two techniques for estimating the trust relationships between the entities:

1. Direct Trust derives from direct interactions among the entities, in order to estimate the trust level of the trusted entity. A number of factors influence the (re)evaluation of the direct trust in a certain entity:
   - The history of their interactions which includes the number of previous interactions, the frequency of interactions and their duration.
   - Trust ageing, i.e. (re)evaluation of trust based on the time elapsed since the last interaction.

2. Indirect Trust can be applied in cases where there is no direct trust relation between the entities, or the history of direct interactions is limited. Based on the trust levels assessed by fellow Community members, the trusting entity is able to infer trust (cf. referral trust).

Applied Scenario
The Trust Management & Evaluation framework provides the necessary infrastructure for the maintenance and management of dynamically changing trustworthiness of stakeholders with respect to different domains of collaboration. For example, Hans wants to use a service from a classmate called Louise. Hans knows Louise quite well so has a high level of trust in her personally. This is shown as trust relationship R1 in the diagram. However, the "Timetable" service Louise is running was supplied by Google. Thus, Hans trust relationship now includes Louise and Google, shown as trust relationship R2 in the diagram. Finally, Louise is a member of the "Adventure club", Hans is currently not. Hans is trying to work out if he can trust this "Adventure club" community before joining it. This is an example, of indirect trust and is shown as trust relationship R3.
Privacy

Helps an individual to manage their personal data and its disclosure.

A Privacy Policy is a legal document that describes some or all of the ways a party gathers, uses, discloses and manages an end-user’s data. It contains a list of data that can be recorded; why the data is recorded; how that data is stored and in what circumstances is it disclosed.

Innovation area
The ‘privacy by design’ approach of SOCIETIES means that privacy protection is intimately integrated in the platform architecture and not added as “la cerise sur le gâteau” (the cherry on the cake) or as an afterthought.

Privacy protection helps the end-user in the management of their personal data by enforcing:

1. A negotiation mechanism between User Preferences and Service / Community Policies.
2. Data disclosure with personalised data obfuscation mechanisms.
4. Privacy assessment.

Innovation: Personalised Privacy Policy negotiation
Privacy policy negotiation provides end-users with flexibility to choose the personal data they wish to disclose to services, communities or the public. Using preferences to automate the privacy policy negotiation process benefits the end-user in many ways:

1. The end-user is alleviated from the burden of configuring the details of their privacy policy manually.
2. The data owner and the data requester are bound by the terms of the mutual data disclosure agreement.
3. It helps the end-user to be consistent with the data they disclose as the same preference can be applied in more than one situation.

For the end user, a given community’s privacy policy can be used to assess whether the community applies privacy practices which the user approves of. This can be a deciding factor on joining an existing community (CIS). When connecting to an individual (CSS) or a community (CIS), privacy policy negotiation aids the user in tailoring data disclosure to their needs and preferences, i.e. only disclose information that is necessary, or the end-user is happy to share.

Applied Scenario
The following example depicts a scenario where Jack wants to use Abby’s travel agent service. Abby replies with a privacy policy, which Jack examines and evaluates. He then suggests some terms and conditions to Abby, who can accept these terms and conditions or make a compromise proposal back.
to Jack. When Jack and Abby, have agreed on the terms and conditions, an agreed privacy policy is exchanged.

**Innovation: Data obfuscation**

The idea behind data obfuscation is to reduce the personal content of data exchanged. This technique results in two privacy by design principles: data minimization (share only necessary data) and enforcement (be sure that data are well protected), being satisfied. On connecting with other individuals or communities, the end-user’s data will be obfuscated to the desired level before being shared with others.

Location and identity have already been obfuscated in other projects. In SOCIETIES, obfuscation is formalized, generalized and applied to many types of data, e.g. location, name, age, nationality, etc.

**Applied Scenario**

The following diagram gives a simple example. Anne wishes to use the “Recommended Reading” service provided by the ACME company. This provides book recommendations to people who sign up. The service requests access to Anne’s location from Anne’s context broker. Anne’s context broker checks the privacy policy and discovers that this attribute needs to be obfuscated. The Context broker forwards the request and current data value (“75002 Paris, France”) to Anne’s obfuscation manager. Based on the agreed privacy preferences, a level of obfuscation is given as 50% and this is applied. The location of Anne is reported to the ACME “Recommended reading” service as “France”, so the service shows books available in France.

**Innovation:**

**Personalisation and learning for privacy**

By monitoring the end-user’s behaviour with regards to data disclosure, privacy preferences can be learnt that can be used to automate the detailed processes of the privacy components such as negotiating privacy policies, selecting an appropriate identity to represent the end-user, obfuscating personal information such as age, location, activity etc and controlling access to personal data.

On interacting with other individual’s (CSS’s) or community’s (CIS’s) privacy plays an important role as it is all about the end-user having control over what exact information is shared.

By combining learning and personalisation, a privacy preference model can be created to represent the preferences of the end-user regarding the use of their data. These preferences are context dependent and can be evaluated on demand to provide a decision on what to do when some entity requests access to some data. Using privacy preferences and a service privacy policy, the system can engage in a semi-automatic privacy policy negotiation process.

The purpose of the privacy policy negotiation is to counter the “take it or leave it” approach that the majority of web services offer today and provide the end-user with a mechanism that allows them to negotiate over the terms of data disclosure and data processing.

The automated learning of privacy preferences facilitates the convenient use of the privacy protection system for the end-user. Where possible most of the negotiation actions are automated, leaving the users work out the exceptions to the rules, rather than all the rules. This innovation makes the privacy protection aspect more usable and convenient. Thus, it is more likely to be effective.

**Applied Scenario**

For example, Katia is an inter-railing back-packer and wants to travel round Europe to see the sights. In every city where she stops, she likes to meet fellow back-packers to socialise and enjoy the sights together. She has already travelled from London to Paris to Brussels. From her stay in London, SOCIETIES has applied a set of community privacy preferences from her fellow London back-packers via the “See-the-sights” community. They provided some nice settings for disclosing her location (only the area of the city is disclosed), and to share a list of “sights” with other community members. She shared her food preferences in France, she likes Italian, but every other night, she prefers to sample the local cuisine. The SOCIETIES system has picked up on this trend and automatically learnt the corresponding privacy preferences. When she arrives in Amsterdam, the system automatically negotiates with the local “See the sights” community. This gives her a list of other backpackers who also want to see some the same sights, and maybe they can get a group discount on “Anne Frank’s” house.
Community Context

Managing context for dynamic communities of end-users

Innovation area
From one perspective, a community can be seen as an individual entity with its own community attributes that are derived from the individual community member’s attributes. However, not all the individuals’ attributes are meaningful to be “translated” into community attributes. For example, there is no need to extract community context from the email address or the name of an individual.

To our knowledge no context models have been proposed that can support community context management in large scale systems. Some research has been carried out regarding community context models but there is no provision for conflict-types or inheritance. Some research has been done on forming groups of people with common interests but management (e.g. update) of these groups is not performed based on end-users’ context changes, e.g. change of location or activity.

Innovation: Context modelling
The community context model assists with community management, helping people in both the digital and physical world to discover people, services and devices that are of interest. It allows the modelling of context information for:

1. communities of individuals
2. quality of context
3. context history
4. social relationships
5. bonds among end-users
6. interactions among end-users

These new context types for community/group context information also generate new conflict scenarios when this information is combined. These new conflict scenarios require new solutions (e.g. inheritance procedures, and conflict-types) to resolve them. They can however, in turn, support more complicated and effective prediction algorithms.

Applied Scenario
An example of a new context type for a community, is the location attribute. Suppose that there is a “Jones family” community in Frankfurt airport travelling to Lisbon. There is a “Travel together” service that monitors and compares the community’s current location with each individual’s current location. This service can easily send alerts to an individual if it detects that this individual is located in a restaurant on the wrong level of the airport than the one that the rest of the community is in.

Innovation: Community Context Estimation
Community Context Estimation is responsible for calculating the attribute values for a given community, from the attribute values of the individuals participating in this community. The community’s attributes can change more often than the attributes of the individuals belonging to this community. For example, each one of the members of a community has an age which changes once in a year, while the age of the community changes more often as new members join the community or existing members leave the community.

The Community Context Estimation supports both on-demand, as well as continuous, community context estimation. This can provide useful real-time information for community activities and connect people with common interests.

Various models can be used to represent and estimate the Community Context values, such as:

- Aggregation of the values of the context information of individual community members
- Stochastic representation of the above (mainly for the discrete or enumerated context value formats)
- Average values (for the discrete or continuous context value formats)
- Median values (for the discrete or enumerated context value formats)
- Most probable value, etc.

Applied Scenario
Luigi is a doctor, and is attending a conference in Rome for heart specialists. The “Da Vinci” hotel has a wall display located beside the coffee station. Luigi along with his fellow doctors joined the “Mio cuore” community offered by the Medical Association at the conference. The hotel wall display starts showing a video advertisement for medical covers for an I-Pad.

Knowing some of the community’s attributes such as age, profession, location and so on, a service provider can run more targeted advertisement campaigns, such as, presenting in a specific room through a display device (e.g. digital signage) content that it is addressed to the specific community’s estimated attributes.

Innovation: Context Similarity Evaluation
Context Similarity Evaluation (CSE) estimates the similarity of a set of context information within a set of entities. This is used to see what individuals have in common, or to spot differences e.g. calendar scheduling. CSE is a new research area which requires new algorithms/mechanisms for context comparison. Quantifiable context information such as location co-ordinates, weight or temperature will require arithmetic mechanisms to evaluate the similarity of context. Qualitative context information such as end-user interests, status or symbolic location will require new principles in comparing context semantics to estimate similarity.

Context Similarity Evaluation (CSE) estimates the similarity of context information which will be used by other components to aid in decision making for some or all of the following tasks:

1. Creation/discovery and deletion of communities;
2. Identification of sub-communities and of communities that should/could be merged;
3. Community membership management (i.e. addition/removal of members);
4. User intent prediction based on similar context (i.e. similar context may indicate the same intentions);
5. Preferences discovery based on similar context (i.e. the same preferences may be applicable and may need to be triggered not only under a given combination of context information values, but also when similar context conditions are observed).

Applied Scenario
The conference application for the “Testing complex system’s” event, asks that everyone provides some profile information on what they have used in the past to test IT systems. It also allows them to express interest in topics they would like to know more about. Eric has already downloaded the application and filled in the registration information. At the conference, the conference service offers suggestions, for other attendees he might like to meet. Behind the scenes the conference service is using context-similarity to compare people’s professional profiles, and where there is sufficient similarity, highlighting this to the end-user.
Context

Innovation area

End-user locations, are a real differentiator in ubiquitous computing, and enable a large set of applications, services, business processes, analytics and business intelligence, transferring some abilities from the web world to the mobile physical world. Within SOCIETIES, the physical location of the end-user is captured as a context attribute; with special algorithms, software and hardware used to capture and improve the precision of information.

Innovation:
Location Prediction by Individual and Collective Behaviours

Being able to predict the end-users location in advance allows the system to take proactive steps to ensure that the system and third party services are ready for the impending location change rather than just reacting after the change is made.

Within SOCIETIES, the granularity of location prediction is at the cellular tower level, since, it is focused more on predicting end-user’s activity over a long time span. This is complementary to indoor location inference innovation. While the state-of-the-art research predicts the mobility pattern of individual end-users based on their historical location traces, we propose to use the collective behaviour of massive end-user groups mined from cellular trajectories to enhance the prediction of individual end-user’s future locations. This leads to a new type of location predictor considering both individual's historical mobility pattern and the collective behavioural pattern of the end-user communities. The evaluation results show significant improvement in end-user location prediction. The proposed approach differs largely from other existing solutions.

Our predictor collects the historical location records for a set of end-users, and it predicts the locations of each user for the next 6 hours. Innovatively we view the large population as a whole system, and we believe individuals’ mobilities are correlated with each other. To the best of our knowledge, we are the first to introduce such a viewpoint derived from newly-fashioned complex human dynamics to the end-user location prediction.

To investigate the correlation between each user’s mobilities, we treat all end-users’ locations at the same time slot as a transaction. Then we use association pattern mining techniques to extract the high confidence rules. The figure above shows the association rules we mined from 5 months continuous tracking (we adopt MIT’s reality dataset here to

Improving the precision of indoor location, and unifying outdoor location to provide rich context.
evaluate our predictor. It shows many high confidence rules exist. These association patterns are what we called Collective Behavioural Patterns which identify how end-users’ mobilities are correlated.

To adopt Collective Behavioural Pattern for location prediction, we implement a Naïve Bayes based Collective Behavioural Pattern Learner (CBP) to learn and predict an individual’s locations in the next 6 hours from the current locations of other end-users. However CBP can achieve an accuracy of around 48%, it is still hard to predict all locations by adopting collective behavioural patterns, since generally human mobilities are independent. Therefore, we boost our CBP-based solution with the Markov-based individual mobility predictor to form a Hybrid predictor. The evaluation shows that the Hybrid predictor outperforms both CBP-based and Markov-based solutions significantly.

**Applied Scenario**

Nicole is a student. Every Thursday she attends school, and in the afternoon tries to do some sport. On days when it rains she prefers to study, on dryer days she likes to play football with some of her classmates.

The SOCIETIES location prediction mechanism presents the temporal-spatial/collective-behavioural regularity of personal movement from the current location to a future one. The location(s) of an end-user in the next 6 hours strongly represents what she wants to do in the near future due to the specific semantics of each location. For example, being located in the “Football pitch” suggests an intent of partaking in sport; while being in the “library” implies that she intends to study.

**Innovation: Indoor location**

The concept of Geo-social networking, which provides location based capabilities to enhance the social dynamic experience, is becoming widely accepted. It usually depends on end-users submitting their indoor or outdoor location. SOCIETIES Presence zones will enhance this experience by estimating with fine granularity the indoor contextual location of end-users in a completely seamless manner. The information collected by the system is used as a context source to provide better analytics, recommendation, personalisation and user intent prediction. The state of the art solution differs from other commercial solutions in the following aspects:

1. **Accuracy and granularity** – Sophisticated, patent pending methods are used to calculate the indoor location of entities in the monitored environment. Environmental metrics are collected and aggregated using machine learning techniques. The system works in real time and supports up to ten thousand location updates per minute.

2. **Seamless flow** – In “real world” scenarios end-users are reluctant to install third party software on their mobile devices that reads their personal information, uses data from their data plans while connecting to the network and consumes valuable system resources while doing all sorts of calculations. Presence Zones is seamless in that sense. No application has to be installed on the mobile device and no “check-ins” are required.

3. **Inexpensive hardware** – We use WIFI equipped sensors to monitor the source of WIFI communication in the environment, without of course accessing in any way the content of the communications. The hardware for the sensors is standard and inexpensive, allowing off the shelf thin clients.

4. **Environment independent** – The solution can be deployed...
in any environment. No integration with access points or any other infrastructure in the environment is required, and there is no dependency on any external service provider, such as an ISP. It is an almost plug & play solution that can easily be adapted anywhere. The only requirement is that WIFI be installed in the venue.

In relation to equivalent systems, Facebook, Foursquare and Google latitude all require some form of check in. Google’s indoor navigation solution requires installing software on the mobile device. Other commercial products either utilize high-end RF sensors or perform sensing from the WIFI access point. In the latter the accuracy becomes a function of the number of access points and thus many access points have to be deployed.

**Applied Scenario**

Anne is a student at Heriot-Watt University. She enters the lecture hall to attend a class. Afterwards she leaves the class to get some lunch. Both events are detected by the Zone Sensor WIFI access point and forwarded to the Presence Zones service back-end. This back-end exposes this information as a context source. The Context inference engine then aggregates the information to infer a new fact, that Anne is in Heriot-Watt University. Further analytic processes can be run on the data to support visualisation and prediction activities. For example, predicting where Anne and her classmates will be next Tuesday.
Contact Information

Project Coordinator:
Kevin Doolin
kdoolin@tssg.org

Technical Coordinator:
David McKitterick
davidx.mckitterick@intel.com

Research Coordinators:
Micheal Crotty
mcrotty@tssg.org

Nick Taylor
N.K.Taylor@hw.ac.uk

Project website:
http://www.ict-societies.eu/

Twitter:
http://twitter.com/ict_societies

Facebook Page:
http://www.facebook.com/ICTSOCIETIES

Open Source Page:
https://github.com/societies

Deliverables:
http://www.ict-societies.eu/project-deliverables/