Acronym: ProACT
Title: Integrated Technology Ecosystems for ProACTive Patient Centred Care
Call: PHC25-2015: Advancing ICT Systems and Services for Integrated Care
Duration: 42 Months
Website: www.proact2020.eu
Grant Agreement No.: 689996

Document Identification				
Deliverable ID	**D3.9**	Deliverable Title	**A set of person-centred analytical methods for risks and outcomes (Update to D3.2)**	
Release	Version	2.0	Date	31/10/2017

Key Information	
Deliverable description	Report on the development of a set of person-centred analytical methods for risks and outcomes. The initial report was delivered in month 10 with updates due in month 22 and 39.
Deliverable type	Report
Original due date (month number/date)	M22 (30 October 2017)
Real due date (month number/date)	M22 (30 October 2017)
Principal Author (name/entity)	Julia Lastra (TREE)
Partners Contributing (name/entity)	Jaime Medina (TREE), Julie Doyle (DkIT), Alessandra Pascale (IBM), Emma Murphy (TCD)
Internal Reviewer (name/entity)	Asel Kadyrbaeva (EASPD) and Lorenzo Desideri (AIAS)

Dissemination Level	
Restricted	
Public	X

Release History

Version	Date	Status*	Revision Comments	Author
V0.1	14/09/17	D	First draft version	Julia Lastra
V1.1	27/09/17	C	First completed version	Julia Lastra
V1.2	28/09/17	C	Second completed version internally reviewed by TREE	Jaime Medina
V1.3	04/10/2017	C	Updated version reviewed by DkIT	Julie Doyle
V1.4	05/10/2017	C	Updated version reviewed by TCD	Emma Murphy
V1.5	06/10/2017	C	Updated version reviewed by IBM	Alessandra Pascale
V2.0	16/10/2017	C	Internal review comments and amendments by EASPD and AIAS	Asel Kadyrbaeva and Lorenzo Desideri
V2.1	25/10/2017	C	Final sign off	John Dinsmore
V2.2	27/10/2017	C	Submitted version	Julia Lastra, Caoimhe Hannigan

*Status of deliverables is indicated by abbreviations/terms as follows:

Draft (D): The deliverable is partially complete or complete but under review/revision before release.

Complete (C): The final deliverable document is 100% completed, reviewed and authorised for release by the partner responsible for the deliverable or the WP leader.

Revised (R): The final released document has been modified/updated with new content.

D3.2b A set of person-centred analytical methods
for risks and outcomes
ProACT

Table of Contents

Acronyms and Abbreviations

ADL	Activity of Daily Life
API	Application Programming Interface
BCW	Behavioural Change Wheel
BP	Blood Pressure
bpm	Beats per minute
BPPS	Bio-Physical-Psycho-Social
CA	CareAnalytic
CHD	Chronic Heart Disease
CHF	Congestive Heart Failure
DIA	Diabetes
FT	Friendly Trial
HCP	Healthcare professional
mmHg	Millimetre of mercury
MT	Main Trial
PoC	Proof of Concept
PwM	Person with Multimorbidity
QoL	Quality of Life
RS	Recommender System
UE	User Engagement
UX	User Experience
WP	Work Package

D3.2b A set of person-centred analytical methods for risks and outcomes

ProACT

Executive Summary

The overall aim of ProACT is to develop an open web API ecosystem to integrate a wide variety of new and existing technologies to pull, aggregate and analyse data for the purposes of higher order inference, and to improve and advance integrated care for PwM (people with multimorbidity) (including associated comorbidities). The ecosystem will connect four key care and support models central to understanding and implementing effective, continued and coordinated PwM centric care (including self-management). These models are: 1) homecare (including informal care) 2) hospital care 3) community and social care and 4) social support networks.

The developed system, InterACT, will include several *CareAnalytics*, that is, algorithms that process user data to produce information. This information can be displayed back to CareApps (tablet/web-based applications accessed by the PwM and support actors) or can be exported in a convenient format such that the ProACT team can have an analytical system's overview

As described in the grant agreement, this deliverable is an update of D3.2 (M10) and will have another update in M39. The first version of this deliverable focused on conceptual descriptions of *CareAnalytics*, explaining the main ideas behind them and introducing the core machine learning techniques. This deliverable is the result of adding data availability conditions and hypotheses to the first *CareAnalytics* catalogue in order to obtain, as an iterative process, the second *CareAnalytics* catalogue. In fact, this catalogue may vary during or after the Proof of Concept (PoC) trial, since it needs to be implemented on real data, from which it characteristics will be determined.

The document is structured as follows: Section 1 contains a brief description of the first *CareAnalytics* catalogue to provide a general overview of the underlying context. In Section 2, some data characteristics, both from the Friendly Trial (FT) and from the data to be gathered during the PoC trial, are described. Finally, in Section 3, the updated *CareAnalytics* catalogue is introduced.

1 Introduction

One of the objectives of the person-centred analytics is to provide a holistic view of the status of a person based on the BPPS (biological, physiological, psychological and social conditions) model. According to the 4 key care support models in ProACT, the information extracted from the *CareAnalytics* will be oriented to the person himself/herself, to the informal carer, to the formal carer or to the healthcare professional (HCP).

The output of the *CareAnalytics* translates not only into an immediate result of an analysis displayed in a *CareApp*, but it can also translate into an intermediate process that, by using *machine learning* techniques, allows to improve the performance of the general system. Such intermediate processes allow, for example, to add intelligence to the system or to build analyses to obtain key metrics for a better functioning of the ecosystem. The results of these types of processes are not shown directly to any of the 4 key care support actors, but they have an ultimate impact on their user experiences and consequently on the quality of life (QoL) of the patient.

Advances made in other WPs (especially in WP2, WP3 and WP4) have allowed us to redefine the initial *CareAnalytics* catalogue. The updated catalogue does not introduce a new or different approach. It is rather a result of an iterative process that allows further refinement as we reach different ProACT milestones, such as the FT, the PoC trial or the *CareApp* development.

In fact, over the coming months there will be more iterations that will help to further refine the analytical methods during and after the different action research cycles that compose the PoC trial. The iterative nature of this process gives stability to the methods used in an evidence-based way, thus enabling the creation of optimal analytics tailored to ProACT. The process and the proposed timeline are summarized in Figure 1.
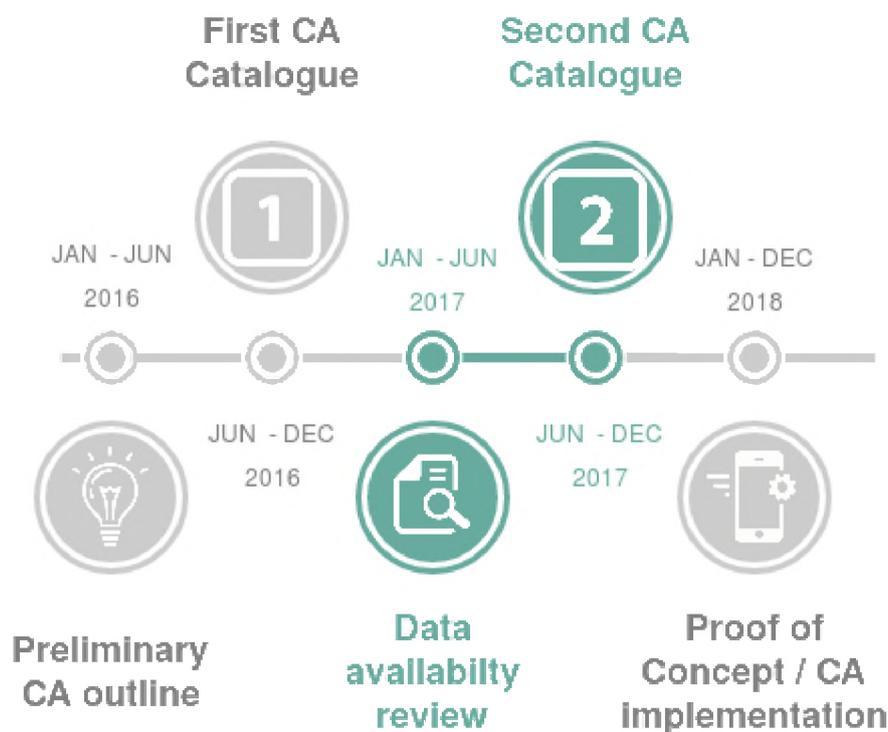
Figure 1: CareAnalytics (CA) iterative process and evolution

This document explains the transition from the first *CareAnalytics* catalogue (outlined in the initial deliverable D3.2 *A set of person-centred analytical methods for risks and outcomes*, M10) to the second one, as a consequence of the results of the data availability review as well as the progress in research and in other WPs.

A summary of the first *CareAnalytics* catalogue is outlined below. For a more detailed explanation see the previous version of this document (D3.2).

1.1 First CareAnalytics approach

Having in mind the needs and interests of the PwMs, the following 9 *CareAnalytics* were initially suggested:

CA 1 **Measurement evaluation and risk detection**: evaluation of the readings to detect problems with the sensors (outlier detection) and measurement classification in terms of risk.

CA 2 **Finding problems in filling out questionnaires**: find out, by using pattern recognition techniques, which questionnaires or questions are more difficult for the PwM to complete.

CA 3 **PwM clustering**: find different groups among PwMs to extract common usage and reporting features from them.

CA 4 **Recommendations on self-management targets**: automatic goal recommendations.

CA 5 **Automatic determination of the self-management behavioural change stages**: measuring how close the self-management goals and/or behavioural change targets are to being achieved by looking into the data previously recorded.

CA 6 **Best function interventions**: calculate a way of finding the best intervention functions (e.g. education, training, persuasion) or key recommendations (education, compliance and sustainability of self-management and care) to improve the outcomes for the PwM.

CA 7 **Probability of behavioural change in a group of PwMs:** locate which PwMs are finding it difficult or easy to manage their conditions using ProACT.

CA 8 **Common intervention functions useful for different groups of PwMs:** define a set of common intervention functions useful to various sub-groups within the ProACT cohort (e.g. females aged between 75-85).

CA 9 **How groups of PwMs interact with the application:** find out how groups of PwMs interact with the application (patterns) in order to assess their engagement. This analytic will be a novel application developed by ProACT.

This catalogue of *CareAnalytics* was the initial starting point of the iterative process described above. At the time when this initial catalogue was defined, there were almost no data availability conditions. In the next section, the first conditions on data are explained as part of the iterative process of the *CareAnalytics* definition.

2 Data review

This section provides an overview of the two sets of data used for updating the *CareAnalytics* catalogue: data from the FT (section 2.2.1) and data gathered during the PoC trials (section 2.2.2). All updates also consider progress in other WPs and in the main *CareApp* development.

2.1 Friendly Trial data

A total of 10 participants from Ireland and Belgium have taken part in the FT. The participants were members of the ProACT research teams at each trial location who tested the platform's performance for approximately three months between M13-16. The data was collected via ProACT system and stored in CABIE and the InterACT cloud.

This section is focused on the values and statistics of the data. For a more detailed description of the actions involved in the FT, see D5.1 *Evaluation Report on Outcomes from the Friendly Trial*.

The variables measured in the FT data are summarized in Table 1:

Table 1: Variables measured during the Friendly Trial

Variable	Description	Units	Provider
Type: *physical/status parameters*			
Blood pressure	Diastolic and systolic values	mmHg	Manual
Pulse	Pulse	bpm	Manual
Blood glucose	Blood glucose	mmol/L	Device + Manual
SpO2	Blood oxygen level	%	Device + Manual
Weight	Weight	Kg	Device/manual
Type: *activity*			
Distance	Daily meters walked	meter	Device
Calories	Daily calories burned	Calories	Device
Steps	Daily step count	steps	Device
Sleep stages	Time of light, deep and REM stages	secs	Device
Time awake	Time in bed awake and asleep	secs	Device

Type: *location*			
PIR Event Timeline	Active or inactive timeline	location	Device
Contact sensor timeline	Opened or closed timeline (doors)	location	Device
Location timeline	Location timeline	location	Device
Temperature timeline	Temperature in the location	Celsius	Device
Presence timeline	Presence in home (present / away)	state	Device
Type: *self-report*			
Mood and anxiety	Self-report scores	%	Self-report
Pain level	Self-report scores	%	Self-report
Sleep	Self-report scores	%	Self-report
Social	Self-report scores	%	Self-report
COPD: Breathlessness	Self-report scores	%	Self-report
COPD: Sputum	Self-report scores	%	Self-report

2.1.1 Involvement

All the variables measured in the 10 participant-sample have produced a total amount of 820,000 entries, with the input timeline shown in Figure 2:
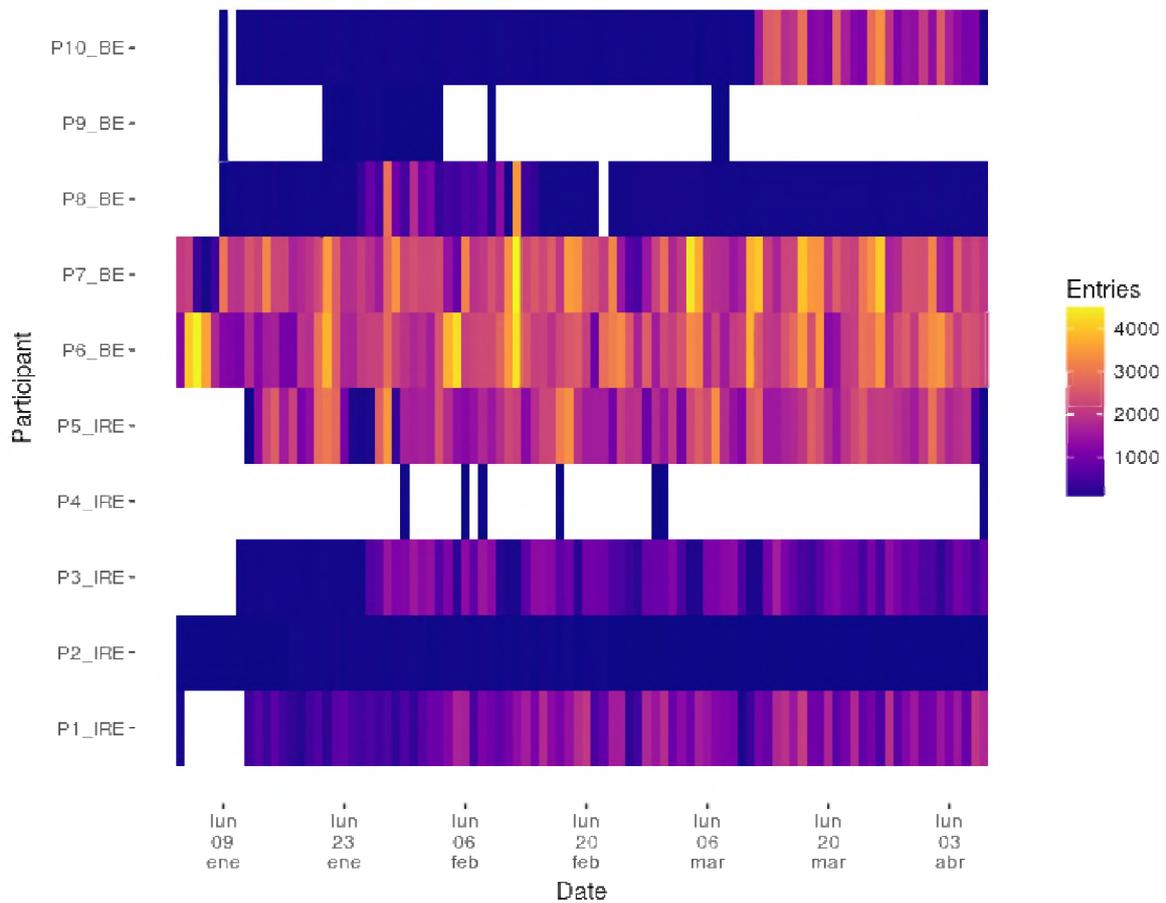
Figure 2: Total of entries per participant and per day during the Friendly Trial

We distinguish between two types of data here:
- data that goes directly to the system without any interaction of the user (data from location devices), and
- data that needs manual interaction to be saved (not only data manually entered, but data coming from other devices that need to be turned on by the user or that needs to be worn).

The second type of data (manual interaction) represents the 2.5% of the entries during the FT (approx. 20,000 entries). The distribution of the participants and the days is shown in Figure 3.
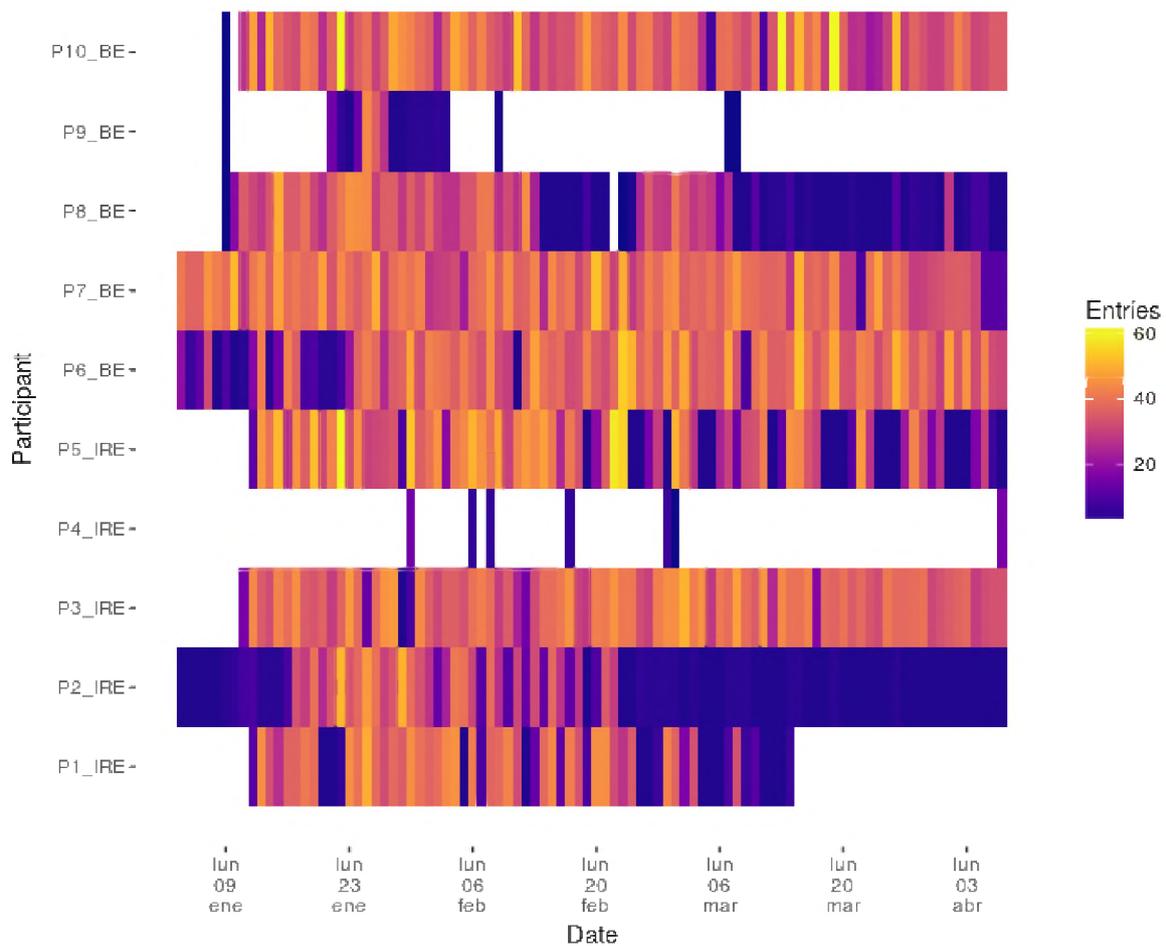
Figure 3: Total of entries (excluding location data) per participant and per day during the Friendly Trial

The number of entries is more balanced when excluding location data. To have a sufficiently large dataset of observations, the user needs to measure his/her significant parameters with certain regularity. These parameters can be employed later on to train algorithms in order to automate processes, to make estimations or to add intelligence to the system. It is important to have a global vision of the usage in order to identify weaknesses, implement improvements and deploy new versions based on the users' behaviour.

Patterns in the time of the day when the *CareApp* is used are analysed to find the optimal moment for sending alerts or reminders. Looking deeper into the data of the FT one can think that, depending on the user, there exists a time interval when it is more likely that the user logs into the system. Figure 1Figure 4 shows how often three participants (P1_IRE, P2_IRE, P8_BE) have been sending data. Each time an entry (observation) was sent to InterACT, a horizontal line is plotted, whose height is given by the sending time. The blue bar shows when data coming from automatic devices is sent to InterACT (smart watches): the storage does not depend on the utilization of the *CareApp*; but data is being sent continuously when doing/finishing some of the activities (e.g. sleep) are measured by the

device. Manual (self-reports) and manual + device (blood pressure, blood glucose, etc.) modes of data entry imply a person has interacted with the system to send information: participant P1_IRE used the *CareApp* between 10:30 and 17:30, never at night nor in the morning. By contrast, P2_IRE used the devices after 20:00 and answered the questionnaires in the morning.

Figure 4: System usage in terms of variables and times for participants P1_IRE, P2_IRE, P8_BE

¡Error! No se encuentra el origen de la referencia. shows the measurements (and requency) that should be gathered according to the simulated role of each participant.

Table 2: Friendly Trial measurement frequency

Participant	Blood pressure	Blood glucose	Weight	SpO2	Self-report
P1_IRE	daily	2 x daily	daily	-	daily
P2_IRE	daily	2 x daily	daily	daily	daily
P3_IRE	daily	-	-	daily	daily

P4_IRE	daily	2 x daily	daily	-	daily
P5_IRE	daily	2 x daily	-	daily	daily
P6_BE	daily	2 x daily	daily	-	daily
P7_BE	daily	2 x daily	daily	daily	daily
P8_BE	daily	2 x daily	daily	-	daily
P9_BE	daily	-	daily	daily	daily
P10_BE	daily	2 x daily	daily	-	daily

With this *roadmap* it should be possible to have a sufficiently large time series. That is, if these parameters were measured with the expected frequency, it would be possible to obtain sufficiently large time series data to implement a deeper analysis or even preliminary *CareAnalytics.* That would be a trivial task if the data collection process was an automatized task. But this is not the case in the ProACT project, since human interaction is behind the most significant data gathering process. As a consequence, *CareAnalytics* will have to deal with uncertainty and incomplete data from their initial stages (design).

The following figure shows the results about measuring the parameters in Table 2Table 2: Friendly Trial measurement frequency the expected frequency:
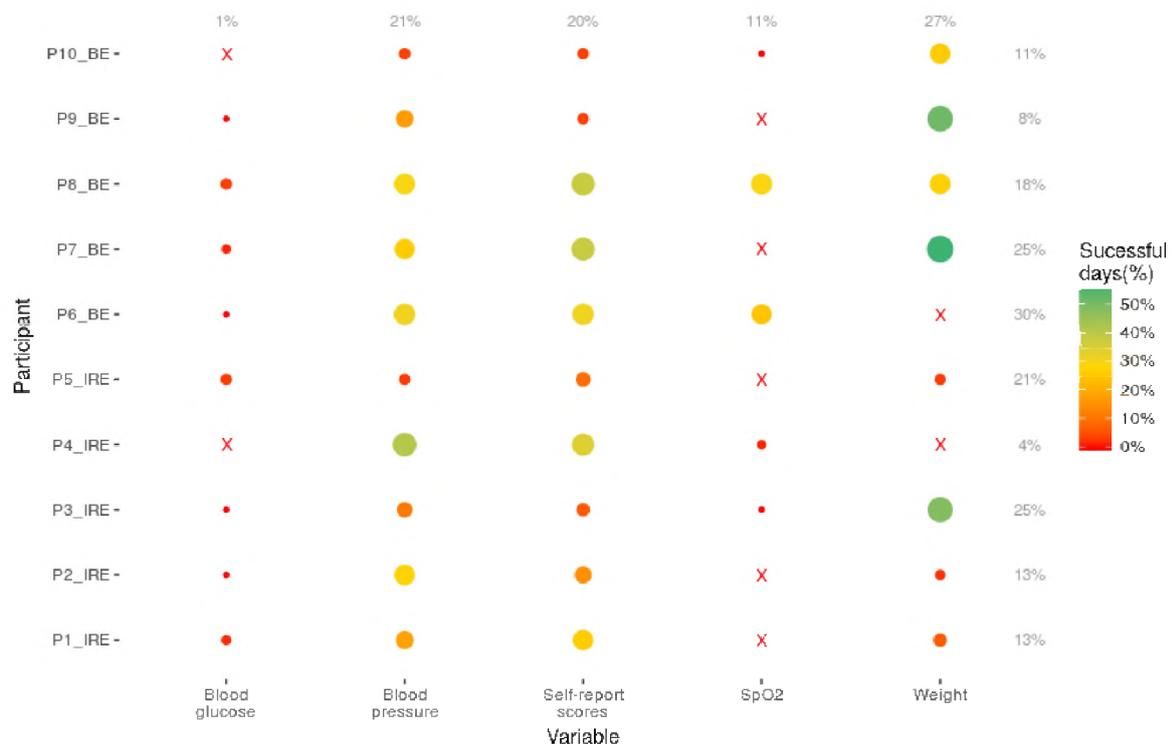
Figure 5: Successful days (in terms of measurements gathering)

Each point represents the percentage of days in which the participant saved the measurements as expected in Table 2: Friendly Trial measurement frequencyTable 2**¡Error! No se encuentra el origen de la referencia.** (which we refer to as, "Successful days"). The

"X" symbol means that the parameter was not significant for the participant and therefore he/she did not have to measure it. The "greener" and/or bigger the point is, the higher is the number of days when he/she saved the value. Results are calculated as a percentage since there was a small difference between the duration of the trial for different participants. The percentages in grey at the top and on the right side of the graph represent the mean of successful days for the variables and for the participants respectively.

Since the objective of the FT was primarily to test the system, most values were entered manually in order to evaluate data transfer processes. They do not represent real values and might be out of a normal range. Therefore, they cannot be used for carrying out deeper analysis of the values and for building a hypothesis. This, however, can be with the PoC data.

2.2 Proof of Concept data

2.2.1 Proof of Concept trials

The PoC trials will be the main data source for the *CareAnalytics*. The sample size of these trials is 120 PwMs and up to 5 additional support actors in the care network with the PwM's permission (up to 360 participants (PwM's plus support actors) at each site (Ireland and Belgium)).

The trial will take place over 12 months: from January 2018 (phased set up begins from December 2017) to December 2018, and there will be 3 research cycles of 3 months each, with a one month reflective redesign phase after each cycle.

2.2.2 Participants

Participating PwMs will be aged 65 years and over and have at least two of the following chronic conditions: COPD, diabetes, Coronary Heart Disease (including Hypertension, Atherosclerosis, Arrhythmia, Angina) and/or Congestive Heart Failure (CHF). They must be capable of providing written informed consent.

Support actor participants will be aged 18 years and over and provide care to a PwM who meets the criteria above mentioned.

2.2.3 Variables

The set of variables gathered during the PoC will be much larger in comparison to the FT. It is possible to distinguish eight data categories, depending on the nature of the variables:

Table 3. Variables gathered during the PoC

Category: demographics	
Age	Profile input (CareApp)

Gender	
Conditions	

Category: vitals

Blood pressure **Heart rate** **Weight** **Blood glucose** **Oxygen Level**	Manual with device (CareApp)
Breathlessness (COPD) **Breathlessness (CHF)** **Sputum** **Oedema**	Self-report score (CareApp)

Category: assessment

Foot care **Cough** **Mood / anxiety** **Falls / fear of falling** **Social interaction** **Activities of Daily Living (ADL)s** **Medication** **Sleep** **Physical activities**	Self-report score (CareApp)

Category: life style

Steps **Time walked** **Distance walked** **Highest elevation** **Sleep stages** **Time inside/outside** **Home location** **PIR (contact / presence sensor)**	Device

Category: home environment	
Temperature	Device

Category: goals	
Goal category	
Time option (day / week)	
Previous goal achievement (% per day)	CareApp (goal section) (to be determined)
Previous recommendations	

Category: education	
Tagged videos	
Tagged tips	To be determined

Category: usage	
First logging date	
Registration date	
First data recorded date (per data type)	To be determined
Videos seen / timestamp	
Petal clicks / timestamp	
Understanding buttons	

The vitals and assessments may depend on the conditions of the PwMs. This will not be a definitive list as it may be updated throughout the trial and will vary according to participant profiles (combination of conditions and self-management routines).

Between November and December 2017 leading into the first action research cycle, variables in **¡Error! No se encuentra el origen de la referencia.** will be revised to add more details (formal definition, range values, units, etc.). In the same way, metrics and variables related to user engagement will be studied and defined in the upcoming months, since sections in the *CareApp* are still in the stages of final development (goals, health tips and users management).

2.3 Conclusions of the review

The application of *machine learning* techniques might be **limited** because of the **sample size** and the lack of historical data. Using simulated data can be a solution until more

historical data is gathered, but for those simulations it is necessary to know the distribution of the different variables.

There is a hint of a **bias** in the data, since the available variables and the frequency of their collection depend on the conditions of the PwMs. This might mean that some techniques such as clustering do not bring any new insights into the set for PwMs. Furthermore, **not having the same variables for the entire sample** makes it more difficult to develop analytics on and for the full set of users.

In conclusion, to address the challenge of having a small set of potentially biased variables *CareAnalytics* should be focused on gathering as much information as possible and on evaluating the quality of such information. *Machine learning* techniques should be applied only on sufficiently large datasets and, in situations where the available data is not sufficient, an intelligent alternative should be in place (case-specific solutions, inference based alternatives, etc.). Given these conditions, the future updates in the *CareAnalytics* catalogue will target a **goal recommender** (fixed to activity)**,** an **education recommender** (to improve involvement), a **user engagement analyser** and a **data cleaner.** Also a **Probabilistic Health and Wellness Profile Builder** will be used to deal with missing data and to model the profiles of PwMs and their environments. These are further discussed in Section 3 below.

3 CareAnalytics catalogue

Advances in other WPs and in the development of the PwM *CareApp* have allowed us to progress in the definition of the *CareAnalytics* before starting the PoC trial, which will be the main data source of the project. That is, it has been possible to advance the definition of the *CareAnalytics* in terms of objectives and scope, but not in terms of algorithms or analytic functions.

WP4 has been crucial to progress in the definition of the *CareAnalytics,* since *Behavioural Change* (BC) is one of the main pillars of the *CareAnalytics* catalogue. Latest advances in BC in ProACT are included in D4.6 *Behavioural change assessment tool kit.* As stated there, some of the intervention functions such as education and training will be delivered using modes of delivery outside of the technology that we will deploy, such as in-person training sessions, paper-based education and training materials. Most intervention functions and BCTs (Behavioural Change Targets) will be delivered directly through the ProACT *CareApp* with support and collaboration from the care network and the trial research team. That means that some of the variables needed for specific *CareAnalytic*s will not be gathered within the system and in consequence, some *CareAnalytics* defined in the first catalogue have to be moved into the background (CA6 *best function interventions*, CA8 *common intervention functions useful for different groups of PwMs*). Again the lack of specific data leaves out of scope the CA2 *finding problems in filling out questionnaires,* since simplicity and flexibility are mandatory in the *CareApp* design, so questionnaires will be short and easy to fill out. Note that it does not mean that difficulties in filling out the questionnaires will not be studied, but just that it will not be studied automatically through the system. It will be possible however to give some recommendations, after analysing the PoC trials data (full dataset), on how such personalised analytics could be created in the future.

Overall, analytics will be required to support the Behaviour Change Wheel (BCW) methodology used in WP4. To achieve this, the analytical component consists of three *CareAnalytics*: the **User Engagement Analyser,** the **Goal Recommender** and the **Education Recommender.** D4.4 *Report specifying the conditions for effective Behavioural Change with ICT-AT in ProACT* contains the update on the behavioural change targets and intervention strategies for the project, which have made these *CareAnalytics* important.

For dealing with the PwM model and their environment, two *CareAnalytics* will be developed: the **Probabilistic Health and Wellness Profile Builder** to exploit the results of the PwM model in order to estimate vulnerability and risk as part of the care management process, and a **data cleaner** that has two main objectives: detect anomalies and outliers in data and establish automatic personal thresholds.

The following table summarize all the relevant information concerning the *CareAnalytics*. It has to be mentioned that the deadlines for each analytic are set again having in mind the time needed to train them, since the *CareAnalytics* need real data (coming from the PoC trial) to be implemented.

Table 4. CareAnalytics in ProACT

CareAnalytic	Input	Output	End user	Estimated deadline	Developed by
User Engagement analyser	Usage data	Coefficient of engagement (%) on the different parts of the system	ProACT team	M29	TREE
Goal Recommender	Demographics, life style, goals data	Recommended value on specific objectives (activity)	PwM	M29	TREE
Education Recommender	Vitals, Usage data, User Engagement Analyser output, Bayesian Analytics output	Type of tip, priority list of educational content	PwM	M33	TREE
Probabilistic Health and Wellness Profile Builder	Vitals, demographics, assessments, life style	Conditional probabilities, prediction, missing data estimation	ProACT team	M29	IBM
Data cleaner	Vitals (numeric), Home environment	Vital thresholds, data classification (normal / outlier), confidence levels	ProACT team	M25	TREE

Note that the input may be a subset of the type of data mentioned; it may occur that not all the variables are significant for the models.

3.1 Goal recommender

A recommender system (RS) is a software tool and technique providing suggestions for items to be of use to a user (Mahmood et al., 2009; Burke, 2007). The suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, what online news to read or in the context of health what educational material may be of benefit to read. "Item" is the general term used to denote what the system recommends to users.

Recommender systems have proven to be a valuable tool for online users to cope with information overload. They have become one of the most powerful and popular tools in electronic commerce. Correspondingly, various techniques for recommendation generation have been proposed during the last decade, many of them have also been successfully deployed in commercial environments (Ricci F. et al, 2011).

Development of a RS is a multi-disciplinary effort which involves experts from various fields such as Artificial intelligence, Human Computer Interaction, Information Technology, Data Mining, Statistics, Adaptive User Interfaces, Decision Support Systems, Marketing or Consumer Behaviour.

Recommendation systems use a number of different technologies, which can be classified into two broad groups:

- *Content-based systems*: examine properties of the items recommended. For instance, if a Netflix[1] user has watched many cowboy movies, then the system recommends a movie classified in the database as having the "cowboy" genre.
- *Collaborative filtering*: systems recommend items based on similarity measures between users and/or items. The items recommended to a user are those preferred by similar users.

One of the most successful technologies for content-based systems, called *collaborative filtering*, has been developed and improved over the past decade to the point where a wide variety of algorithms exists for generating recommendations. Each algorithmic approach has adherents who claim it to be superior for some purpose. Clearly identifying the best algorithm for a given purpose has proven challenging, in part because researchers disagree on which attributes should be measured, and on which metrics should be used for each attribute.

Although most of the applications of RS technologies are focused in Web Content or electronic commerce, new fields of application have been studied. In the context of chronic disease self-management, a few implementations have been made, such as the Integrated Collaborative Filtering framework (Hussein et al., 2014) that provides medical advice for diabetes or the personalised case-based RS for self-management of diabetes during exercise (Chen et al., 2017).

Roadmap

As part of T3.3 (*Person-Centric and Ecosystem Analytics*), a goal recommender will be implemented to assist the PwM using the *CareApp* in choosing a suitable goal. A first version of these analytics will be ready at the beginning of the second research cycle (M29). This first version will work for a specific goal case: activity goals. The objective of the *CareApp* is to provide a suggestion on the number of steps / distance being aware of the user's demographics and his/her past goal activity information (previous recommendations, goal achievements, etc.). Similarity measures (between PwMs) will be studied in order to deal with the *cold start* (recommendations to new users). The definition of validation metrics will allow us to finally test the recommender. As the goal section will not be available in the *CareApp* until the second research cycle (see D2.6 for more details about features introduction), a first version will be developed based on simpler metrics, calculated as a

---

[1] https://www.netflix.com

result of the physical activity made by the PwM. Later versions will take into account his/her past goal activity information.

3.2 Education recommender

Although it is also a recommender, it will work differently than the goal recommender. These *CareAnalytics* will not be based on similarity measures or work as a "traditional" recommendation system, but it will recommend tagged content to improve the engagement of the PwM with the system and therefore, it will try to improve his/her QoL based on increased knowledge acquisition for self-managing.

Roadmap

The Education recommender will facilitate training material (tip of the day, videos or other tagged content) using a smart approach: on the one hand, it will use thresholds on vitals so if any parameter has reached a value out of the normal range (statistical normal range per PwM), the PwM will be presented with specific training material related to that vital. This has to be defined in collaboration with project partners TCD and DkIT, since it may increase the development complexity of the software. On the other hand, this smart approach will use the output of the user engagement analyser to detect a low involvement level in aspects of self-management and will guide the PwM in the usage of the system to improve this. The first version of this *CareAnalytic* will be available in the third research cycle (M33) for two main reasons: some input data is needed to be generated before starting the development of this *CareAnalytic* and it can only be gathered during the PoC trials, and not all the related features will be implemented in the first research cycle (see D2.6). It is also important to map onto the staggered feature introduction process, as part of improving the learning curve for older PwM's adopting a new digital self-management support tool.

3.3 User engagement analyser

User Engagement is a notion used in different fields, not only in a business context but also in a scientific context and standard definition for this concept does not exist. It has been treated as a process itself and also as a state within a process, as an experience, a degree, a connection and as an indicator of the relation between a person and a product.

User Engagement (UE) is related to User Experience (UX): both find it important to appeal to a person's emotions and feelings and to create some form of fidelity to a brand or a product. According to O'Brien et al. (2008), the UE is the quality of users' experiences with technology and is comprised of attention, affect, aesthetics, novelty, interest, control, feedback, challenge, and motivation. This definition is based on conclusions from experiments in four domains: Web searching, online shopping, video games, and educational webcasts (O'Brien & Toms, 2008) and extended the scope of the previous definitions of engagement (Jacques et al., 1995; Webster & Ho, 1997; Skelley et al., 1994). This work also posited that engagement fits within McCarthy and Wright's (2004) 'threads' of

experience framework. According to the compositional thread, engagement is a process whereby users pass through stages of becoming engaged, staying engaged, disengaging, and re-engaging. This process is situated in a specific spatio-temporal context, i.e. time and place. The emotional and sensual threads recognize the affective, aesthetic, and interactive attributes of engagement. These attributes will likely vary in intensity during the experience depending on the interaction between the user, system, and task at a given point in time.

Many of the experiments and surveys concerning the significant factors of the UE have been made on the four domains (Web searching, online shopping, video games, and educational webcasts) mentioned above. These outcomes are the starting point for UE in the project, but they may be updated to deal with specific characteristics of the project: having a skewed PoC trial sample (people older than 65 years that may not be used to technologies), care network intervention, etc. In order for patients to become engaged in device data gathering, key person-centred issues relating to usefulness in care, motivation, the safety and privacy of information, and clinical integration need to be addressed. Because the successful usage of devices and apps requires an ability to comprehend and utilize personal health data, the user experience should account for individual differences in numeracy skills and apply evidence-based behavioural science principles to promote continued engagement.

In order to realize widespread adoption in a real-world community of chronic disease patients, according to Chiauzzi (Chiauzzi et al., 2015), key person-centred questions need to be addressed, namely:

- Is the data useful in care / self-care?
- How can users stay motivated to use devices?
- Does usage lead to behaviour change and affect health outcomes?
- Is the information safe and will privacy be maintained?
- Will providers use this information in treatment planning and delivery?

Again according to Chiauzzi, the following variables are of interest when analysing the UE to digital systems in self-management:
- Good measurement properties of devices data.
- Devices battery life.
- Easier synchronization via Bluetooth and WiFi.
- Inclusion of additional sensor measures.
- Aesthetics from the device and apps (usability and comprehensible data).
- Habit formation (setting cues, routines and rewards).
- Social motivation (sharing or competing for goals with others).
- Goal reinforcement feedback to monitor personal progress.

However, the level of sustained use of devices is ultimately dependent on the disease, patient behaviour, and measurement need.

The behaviour change targets for ProACT have been constructed using the BCW so that they can be easily measured by the system. The first PwM target, to measure and view key

symptoms on a regular basis using the ProACT system, can be tracked by system usage metrics through the *CareApp*. The frequency and times of when symptoms were measured with ProACT or other devices will be recorded as part of the front end system design without analytics. However, the UE analytic can help us to define "what is a regular basis" for individual PwMs engaging with the system in the trial. It is important that we create a clear definition of different levels of engagement, as this will determine a key evaluation metric for ProACT as a behaviour change intervention. PwMs will have individual requirements to measure and monitor their symptoms based on the type and severity of their health conditions. The definition of user engagement needs to be tailored to reflect these individual differences for ProACT PwM users. Additionally, we will create a protocol to determine how the system will react to different levels of engagement. For example, if a participant does not engage with the system for a number of days an alert may be sent through the app for them to receive when they return to the application. For more detail information on this, see D4.6 (*Behavioural change assessment tool kit*) System engagement levels can also be used to personalize or tailor the education content and goal recommendations that a PwM receives. Therefore the output of the user engagement analyser will serve as important metric for behaviour change but also as an input to the education and goal recommender analytics.

Roadmap

As part of T3.3 (*Person-Centric and Ecosystem Analytics*) and in collaboration with TCD and DkIT, an analytical method for the comprehensive evaluation of PwM engagement with the system will be developed. Additionally, a protocol will be created to determine how the system will react to different levels of engagement and its output will serve as input for other *CareAnalytics* such as the education recommender. The first version of this CareAnalytic will be available on the second research cycle (M29).

3.4 Data cleaner

Data cleaning is the process that identifies, corrects and minimizes the impact of errors on results. Data cleaning is emblematic of the historical lower status of data quality issues and has long been viewed as a suspect activity, bordering on data manipulation. Concerns about where to draw the line between data manipulation and responsible data editing are legitimate. Yet all data sources, no matter how well designed and implemented, have to deal with errors from various sources and their effects on results. Although certain aspects of data cleaning such as statistical outlier detection and handling of missing data have received separate attention, the data-cleaning process, as a whole, with all its conceptual, organizational, logistical, managerial, and statistical aspects, has not been described or studied comprehensively.

As it has been observed in the FT data, some outliers and anomalies may be contained (especially concerning the vitals). To deal with these problems and other possible inconsistencies in the numerical values, a *CareAnalytic* will be developed.

Roadmap

The data cleaner will classify the type of data in terms of quality (or reliability) and those analytics that use vitals as input can work on the results of the data cleaner instead of working on raw data if necessary. Additionally, it will calculate the thresholds (normal ranges) for the vitals that will serve as input in the education recommender.

This *CareAnalytic* will be available for the first research cycle (M24). It will be developed within the already existent data, but it is not excluded that an update would be available before the first cycle ends.

3.5 Probabilistic Health and Wellness Profile Builder

The purpose of the **Probabilistic Health and Wellness Profile Builder** is to build a probabilistic model around the PwMs that is able to represent all the variables affecting him/her as well as the conditional relationships existing among these variables. One important consideration is to cover a diversity of inputs including health, psychological and social based data. The mathematical representation used is called Bayesian network (BN).

Overall, this representation allows for the assessment of the most probable state of chosen variables, given the observed state of other variables. For instance, a BN can assist in trying to diagnose a patient presenting certain symptoms by answering queries based on the following areas: "given that this patient has a moderate lung function related to chronic obstructive pulmonary disease (COPD), given his/her medical history, exercise levels, gender and age, how likely is it that he/she suffers from exacerbations of breathlessness?"

Roadmap

This *CareAnalytic* will be partially finished during the first research cycle and it will be available for the second one (M29). For deeper information on these *CareAnalytic*, see D3.1 *A machine-processable representation of the individual and the analytical methods*.

3.6 Data flow description

The *CareAnalytics* presented above are all connected and they will work as a whole data processing system once they will be developed and running. **¡Error! No se encuentra el origen de la referencia.** shows a diagram of data flow for all the *CareAnalytics* developed for ProACT. The diagram is divided in three areas: a) Data Fetching, b) Personal Information (PI), c) Non-Personal Information (NPI). The first section 'Data Fetching' on the left encloses all the data collected and pushed in the monitoring system managed by CABIE. The data are divided into data collected by sensors and data collected by the PwMs using the *CareApp*. Some of them are collected by the specific device but are then recorded in the *CareApp* by the user (marked with dashed blocks in the diagram). All the data collected are then stored in a database dedicated to data containing PI. Those data are then processed in the Health and Wellbeing *CareApp* that feeds back the results from the *CareAnalytics* to the PwMs. The data are anonymized and sent to the right part of the diagram 'Non-Personal Information' where all the *CareAnalytics* that deals with NPI are deployed. Blue arrows in the diagram

mark data exchange happening every time a new data is recorded, red arrows mark data exchange happening at regularly intervals depending on the specific analytic. While PI data are stored on CABIE, the NPI data are stored in InterACT (see D3.4 for more details regarding the architecture of the overall system). The *CareAnalytics* are described before in this section and include the analytics developed by Tree and IBM. The diagram shows the relation among these analytics and how they will cooperate to provide the final outcome to the research team and to the PwMs.
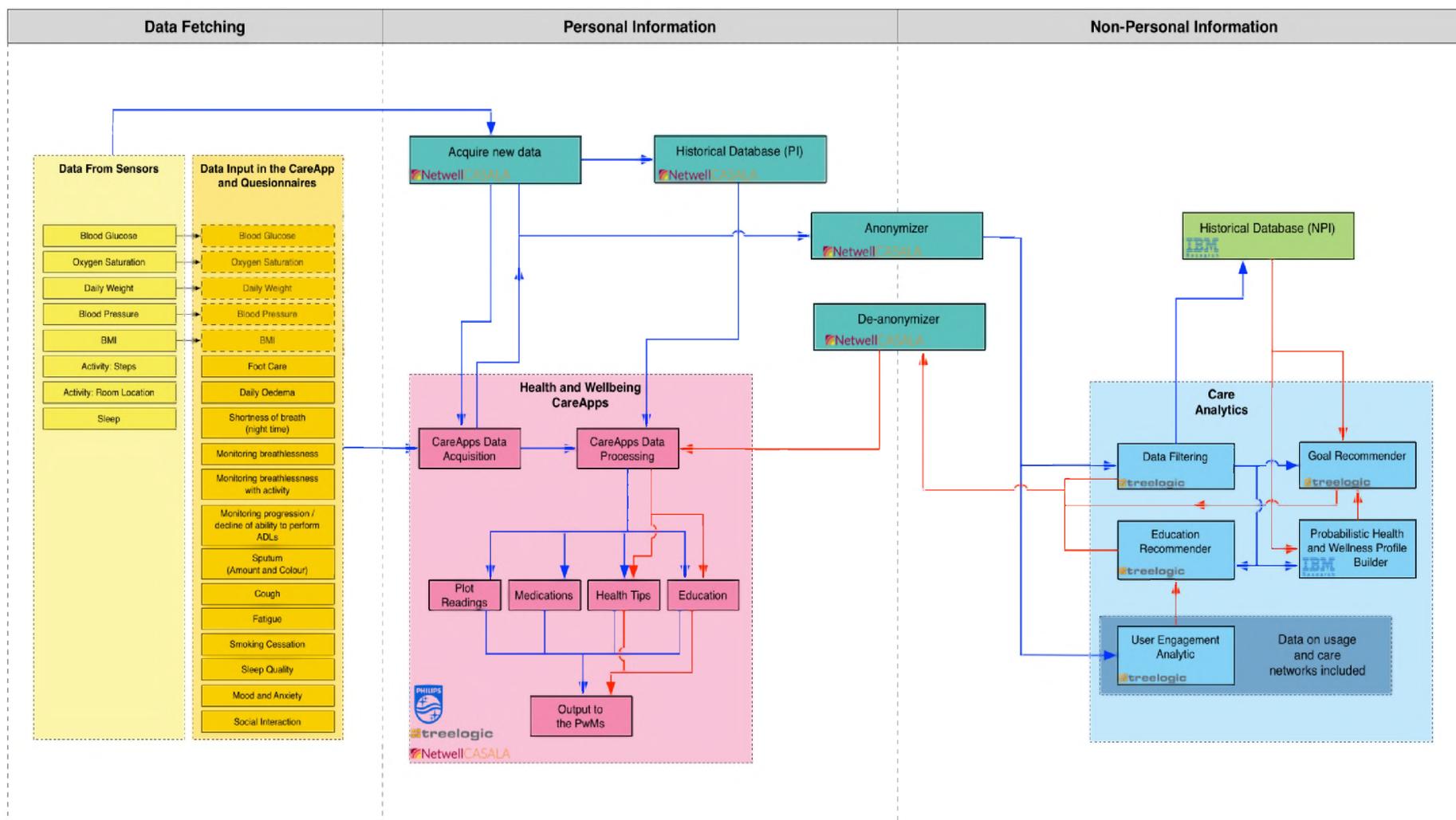
Figure 6. Data flow diagram of the care-analytics developed in WP3.

4 Future work

Next steps concerning CareAnalytics are summarized in the following figure:
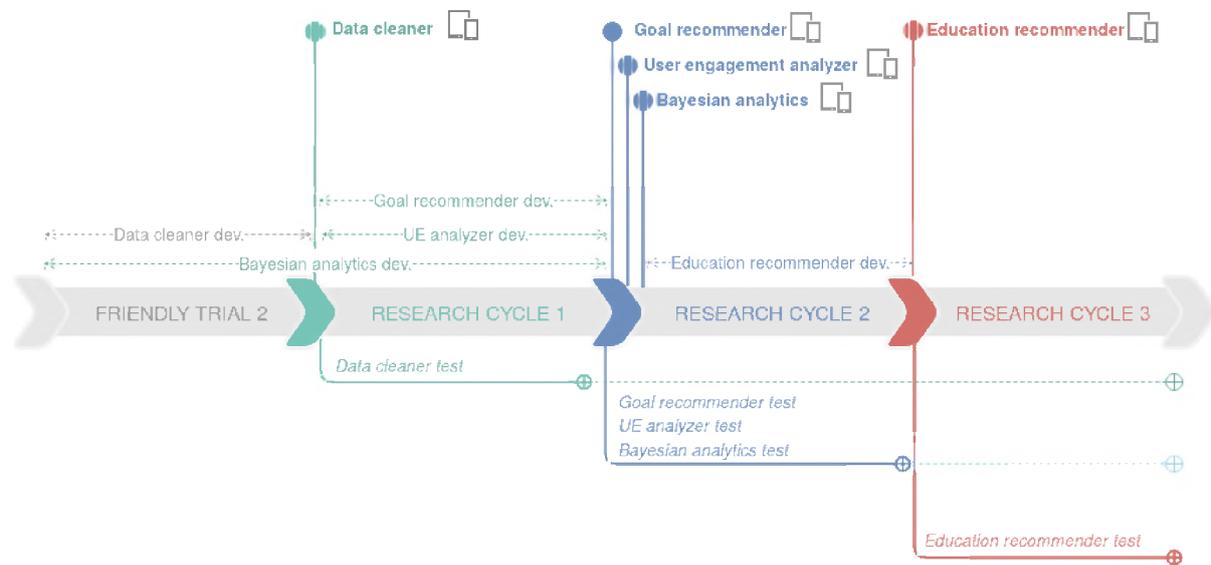
Figure 7. CareAnalytics development process timeline

The *CareAnalytics'* implementation process has always the same structure: a development phase in which algorithms and functions are created, trained and validated, and a testing phase to check its performance (accuracy of the analytical methods and integration in the InterACT system). These stages will be carried out periodically and new releases of the *CareAnalytics* will be launched at the beginning of every action research cycle.

The development of the catalogue will start with the data cleaner and with the training of the Bayesian analytics and will take place during November and December 2017. During this time, the ProACT research team will be testing the final *CareApp* and the devices in order to generate some raw data. With the beginning of the PoC trial, the test of the first analytic will start. At the same time and as soon as specific data (goal, engagement and education) is sent to the InterACT cloud, the development and testing of the respective *CareAnalytics* will take place. Both development and testing will involve the collaboration of other members of the project, since key metrics for user engagement are needed, as well as validation metrics to model the good performance of a *CareAnalytic*. Note that the schedule provided in Figure 7 may be subjected to data availability.

An updated version of this document will be presented in M39. It will cover both the functional and technical description of the *CareAnalytics* (input, algorithms and outputs) and conclusions and remarks obtained from the development process.

5 Bibliography

Burke, R. *Hybrid web recommender systems. The Adaptive Web*, 2007, pp. 377–408. Springer Berlin / Heidelberg.

Chen Y. Y., Wiratunga N., Massie S., Hall J., Stephen K., Croall A., MacMillan J., Murray L., Wilcock G., MacRury S. Designing a Personalised Case-Based Recommender System for Mobile Self-Management of Diabetes During Exercise. *Proceedings of the 2nd International Workshop on Knowledge Discovery in Healthcare Data,* 2017.

Chiauzzi E., Rodarte C., DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. *BMC Medicine* 2015; 13: 77.

Jacques, R., Preece, J., & Carey, T. Engagement as a design concept for multimedia. C*anadian Journal of Educational Communication*, 1995, 24 (1), 49-59.

Hussein A. S., Omar W. M., Li X., Hatem M. A. Smart collaboration framework for managing chronic disease using recommender system. *Health Systems*, 2014, Volume 3, Number 1, Page 12.

McCarthy, J., & Wright, P. *Technology as Experience*, 2004, Cambridge, Mass: MIT Press.

Mahmood, T., Ricci, F. *Improving recommender systems with adaptive conversational strategies*. C. Cattuto, G. Ruffo, F. Menczer (eds.), 2009, Hypertext, pp. 73–82. ACM. O'Brien, H. L., & Toms, E. G. What is user engagement? A conceptual framework for defining user engagement with technology. JASIS&T, 2008, 59 (6).

Maltz, M., *Psycho-Cybernetics*, 1969, New York: Prentice Hall.

O'Brien H. L., Toms E. G., Kelloway E. K., Kelley E. Developing and evaluating a reliable measure of user engagement. Proceedings of the Association for Information Science and Technology, 2008, Volume 45, Issue 1 Pages 1–10

Ricci F., Rockach L., Shapira B., Kantor, P.B. *Recommender Systems Handbook*. Springer, 2011.

Skelly, T. C., Fries, K., Linnett, B., Nass, C., & Reeves, B. Seductive interfaces: Satisfying a mass audience. *Conference on Human Factors in Computing Systems*, Boston, MA, 1994.

Webster, J., & Ho, H. Audience engagement in multimedia presentations. The *DATA BASE for Advances in Information Systems*, 1997, 28 (2), 63-77.

Disclaimer

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 689996. This document reflects the views only of the authors, and the European Union cannot be held responsible for any use which may be made of the information contained therein."