midas

Meaningful Integration of Data, Analytics and Services

Grant Agreement No. 727721

Contract Duration: 40 months (1st November 2016 – 29th February 2020)

This project is funded by
The European Union

H2020-SC1-2016-CNECT
SC1-PM-18-2016 - Big Data Supporting Public Health Policies

Deliverable D5.1 - Review version

Visual analytics tool(s) concept V1

Circulation:	PU
Nature:	R
Version #:	1.2
Issue Date:	17/08/2018
Responsible Partner(s):	Teknologian tutkimuskeskus VTT Oy (VTT)
Author(s):	Juha Pajula, Adil Umer, Markus Ylikerälä; VTT, Gorka Epelde; VICOM, Joao Pita Costa, Flavio Fuart, Dunja Mladenic, Luka Stopar, Marko Grobelnik; QUIN, Simon McLoughlin; IBM, Julius Gomez, Jarmo Pääkkönen, Marika Iivari; UOULU, Oihana Belar, Roberto Bilbao; BIOEF, Paula Kavanagh; HSE, Susan Campbell; BSO (HSCNI)
Status:	Final
Reviewed on:	17/08/2018
Reviewed by:	MIDAS Executive Board

Grant Agreement No: 727721

Contractual Date of Delivery: 17/08/2018

Executive Board Document Sign Off

Role	Partner	Signature	Date
WP1 Lead	Ulster	Michaela Black	16/08/2018
WP2 Lead	SET	Paul Carlin	
WP3 Lead	VICOM	Gorka Epelde	15/08/2018
WP4 Lead	KU Leuven	Gorana Nikolic	17/08/2018
WP5 Lead	VTT	Juha Pajula	16/08/2018
WP6 Lead	DCU	Regina Connolly	17/08/2018
WP7 Lead	Ulster	Jonathan Wallace	16/08/2018
WP8 Lead	Ulster	Michaela Black	16/08/2018

Abstract

This deliverable describes the four pilot sites including stakeholders, scenarios and user stories as well as resources and requirements for visualisation and user experience. The four pilots of MIDAS project are located in Basque Country, Finland, Northern Ireland and Republic of Ireland. The descriptions of the pilot details are addressed in the chapters one to six. The chapter seven includes the identified initial technical solutions to address the needs of the four pilots and chapter eight summarises the identified technologies for the first implementation of MIDAS Dashboard in the first iteration of MIDAS platform.

Copyright

The MIDAS project is funded under the EC Horizon 2020 SC1- PMF-18 Big Data Supporting Public Health Policies

Document History

Version	Issue Date	Stage	Content and Changes
0.1	11.9.2017	Draft	Initial draft based on description of Grant Agreement.
0.2	27.11.2017	Draft	First version with All non-technical details included.
0.3	05.12.2017	Draft	All materials included, submitted for EB review
1.0	13.12.2017	Final	Approved by Executive Board
1.1	14.08.2018	Review	All review changes included, submitted for EB review
1.2	17.08.2018	Final Review	Approved by available members of Executive Board

Statement of Originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Executive Summary

Work Package:	Big Data Visualisation for Public Health Decision Making
Work Package leader:	VTT – Teknologian tutkimuskeskus VTT Oy
Task:	T 5.1 – Design of MIDAS Dashboard for Actionable Information Presentation
Task leader:	VTT – Teknologian tutkimuskeskus VTT Oy

The work of this deliverable is applied within task 5.1 – Design of MIDAS Dashboard for Actionable Information Presentation (Lead: VTT; BSO, DH, SET, BIOEF, VICOM, QUINT). The deliverable defines the four pilots of MIDAS platform for the user interface development including the identified visualisation methods. The four pilots are located at Basque Country, Finland, Northern Ireland and Republic of Ireland.

Key stakeholders have been identified from each pilot and with them the most relevant scenarios and user stories for the pilot topic. The pilot topics are:

- **Basque Country**: Child overweight and obesity prevention prioritised policy
- **Finland**: Preventive mental health and substance abuse of young people
- **Northern Ireland**: Children in Care
- **Republic of Ireland**: “Healthy Ireland” framework and related policies

It was recognised that stakeholders knowledge on technology, use of graphics in support of decision making as well as pilot pilot topics were varying significantly. This led to certain challenges to define a system which could serve them all. To help the work each pilot defined their own key process indicators (KPI) to follow the progress. These were then used to define the WP5 level KPIs.

Specific supporting questions and a living document were defined to collect the critical information from all pilots and this deliverable is now summarising those details. Based on the scenarios and user stories the most critical visualisation methods were identified and made as a target in the first implementation of the MIDAS Dashboard. Other key issues were found in technical side like communication between user interfaces and analytics resources. Applicable approach was found for each of them but implementation of some of them were moved to the second round of iterations.

This document defines the main concept for implementation of the D5.3: MIDAS Dashboard functional prototype V1 and prepares the work related to the deliverable D5.2: Visual analytics tool(s) concept V2. Main end-user interaction will be recorded at deliverable D5.6 Visual analytics tool(s) – MIDAS Dashboard UI evaluation 1.

Table of Contents

1 Introduction

This deliverable defines the used concepts of visualizations and visual analytics tools for different stakeholders in the four Pilots of the MIDAS project. The same MIDAS platform concept is designed to serve all four Pilots. Due to regulatory restrictions each pilot will run independent instances of MIDAS platform with their own policy topics. The core concept of MIDAS is to serve results from large cohorts in graphical visualisations and guide the end-user to use data from large cohorts instead of single cases with a smart user interface (UI). MIDAS Dashboard delivers decision support for policy making and typically policies should be made for large cohorts instead of small minorities. This document defines the minimum requirements of the functional prototype at the first iteration of the visualization concept in the project. The requirements are defined with a twofold approach where the requirements of the visualizations are first identified and then the technical approach is selected so that all identified user requirements are fulfilled. These definitions support also the analysis work of WP4 and data integration of WP3.

The specific methods were defined to select and find the user requirements of each pilot. Each pilot defined their own key process indicators (KPI) from where the KPIs for UI and visualisation concept were defined. Two main tools were used to clarify the requirements and scenarios on each pilot. The first tool was the user personas exercise together with the Co-Design workshop and the second was the living document of Pilot Descriptions. The living document was pre-structured and accompanied with guiding questions to gather a similar level of details from each pilot including:

- Topics of the pilots and scenarios within the topics
- The key stakeholders of each pilot relating to the scenarios
- Identification of existing resources of Policy Board members
- Identification of the needed visualisation and analytics based on current knowledge of key stakeholders

The living document was populated through active participation of each Policy Board partners with the help of user personas exercise, and by continuous collaboration between Policy Board and technical partners. The user personas exercise was initiated to help each of the pilots to understand the needs of the end users. In the user personas exercise each pilot defined average/typical user personas templates to understand what kind of people are involved to the selected research topic. The pilot specific personas templates are found in sections 11.1-11.4 of Appendix 1. These personas were then used by the MIDAS team to generate five general personas for the Co-Design workshop.

At the very beginning of the process, the members of the Policy Board from each Pilot area defined the main policy topic to where the system should provide decision making support and the key stakeholders relating to these. The user personas exercise was then utilized to define the actual needs of potential end users from each of the Pilots. The personas exercise was finished at the Co-Design workshop arranged during the Belfast meeting at fall 2017. Details of the workshop are reported in the Pilot description living document. In the workshop, the user personas developed by the MIDAS team from the pilotwise user personas were studied together with potential use cases from the personas exercise. Also new common user personas were developed collaboratively from the five initial general personas. These common personas are found from Appendix 1 at section 11.5.

Based on the personas exercise the pilots populated the living document, which was then used to define the final requirements and associated details for the MIDAS platform with the technical teams of the MIDAS project. It was agreed that the document will be updated regularly throughout the MIDAS project and thus maintain the description of pilots when the knowledge on selected topics increases and focuses.

Both Policy Board members and Technical teams participated in the Co-Design workshop together with other interested parties. The workshop included also wireframing of the possible user interface for the members of Policy Board with the help of Technical Partners. One of the key actions of the Belfast meeting was to bring the common consensus to all Pilots and in addition to this help the Policy Board Members to define their needs more specifically in the Living document. This also gave a more concrete understanding to the technical teams of what the end-users truly required. The workshop was also utilized to define the cross-site end-user requirements. The final pilot descriptions for the first iteration were defined in the living document with the help of the final user personas clarified in the Co-Design workshop. The responsibility to keep the document content up to date was given to key contact persons of each pilot, the names of these persons were added at the beginning of Living document so that all technical partners can easily find the contact point to each pilot. .

Each pilot defined the process steps how the pilot was applied with Policy Board and reported these in powerpoint slides together with explaining narrative. The state of the process at time of Review is reported at deliverable D6.4 Integrated MIDAS Platform V1 where also the the narratives of each pilot are included. All pilots were looking for options to understand the longitudinal aspect of the identified research question. Most of the pilots were also requesting topics which require geo-locational visualisation for the analytics. All pilots also needed visualisations for visual inspection of data dependencies. The detailed descriptions of the visual requirements of each pilots are described in section 6.

Based on the pilotwise KPIs the global MIDAS KPIs for visualization and UI were defined in reflection to MIDAS DoA objective seven: **" O7: To develop a dashboard for the visualisation of policy models, which will allow policy makers to use visual analytics to assess the impact of changing variables and indicators on their policy decisions (e.g. timeliness, design, assessment, impact). WP5"**
The selected cross-pilot WP5 KPIs are:

- Find and select the key visualisation techniques or types common to pilots
- Find and select the key analytics types to support the policy making, common to all pilots, and the most supporting visualisation techniques for the results of these analyses
- Select the state of the art technologies to implement previous and support the flexibility needed over all pilots at the same time.

As this deliverable is dealing with UI and visualisation related matters the analytics and data related KPI findings are reported on WP3 and WP4 deliverables. From the technical perspective the user personas and the Living document made possible to identify four general main types of users for MIDAS Platform. The user types were defined according to the role and potential way how they would use system. The characteristic user types were:

1. **Decision makers and public health authorities:** end-users who mainly use predefined dashboards. They need to be aware of the current and past situation of the health domain and be able to test the outcomes of possible decision scenarios
2. **Officers, statisticians, clinicians:** the supporting personnel of the first group, who possibly create the dashboards not only for themselves but also for the first group according to the current need.
3. **Researchers:** the healthcare research oriented users, who can even develop the analytics further, define new forecasting models and analytics scenarios.
4. **System management:** updates health system metadata, maintains interfaces to virtual data storage, grants access rights to users etc.

According to the personas exercise most of the MIDAS Dashboard end-users are from types 1 and 2. As earlier described, the Policy board members of each pilot have identified potential users from these groups from the local stakeholders associated with the pilot research topic. The last two user types, the researchers and system management, are using the platform with backend services and within MIDAS project these users are mainly the members of the technical partners. In the future, the group 3 can also include researchers from companies and universities which develops analytics for decision support in health policy making with MIDAS Dashboard.

The technical user requirements were identified with the help of the pilotwise KPIs by identifying the people, environment and potential data involved to the process: Each pilot identified the prioritised policy and key stakeholders for the selected topic. With the help of personas each pilot defined the relevant scenarios and most importantly, user stories which are supported with the MIDAS Dashboard. Each pilot have also identified potential data resources which can support the selected scenarios and user stories. The data resources are described in detail in deliverables of WP3.

The selected policy topics for the four pilots are:

- **Basque Country:** Child overweight and obesity prevention prioritised policy
- **Finland:** Preventive mental health and substance abuse of young people
- **Northern Ireland:** Children in Care
- **Republic of Ireland:** “Healthy Ireland” framework and related policies

The deliverable D3.1 defines the data requirements for these prioritised policies for the research. Detailed description from the data sources and policies are included in D3.1 and this deliverable (D5.1) concentrates on visualisation technologies related to them together with the requirements of UI. Deliverable D4.1 defines corresponding analytics for the identified topics and scenarios.

Based on the pilot definitions the same technical implementation is planned to serve the defined needs. As the same system must serve all of the pilots, the visualization system of MIDAS must support various user stories at the same time. This is gained by developing a flexible dashboard generator where a user can select, add and modify the required analytics and visualizations by themselves and thus create and use the required dashboard. By default the MIDAS Dashboard solution will support most of the typical visualisations for the results, but aach pilot can have also dedicated visualizations, e.g. a map of the target country or special graph used by the local policymakers included in the dashboard via dedicated plugin system. Thus these pilotwise visualisations can be enabled only on the relevant pilot.

The MIDAS Dashboard is built with two main components: UI frontend and OpenVA metadata framework which is build on top of GraphQL signaling framework. The UI frontend is developed with HTML, Bootstrap, JQuery and JavaScript technologies. The UI will have a modular standardized design, which enables any party to develop their own visualization with JavaScript. OpenVA is a Java software developed by VTT and here we utilize the logic model component of the OpenVA framework. OpenVA enables the smart functionality of the dashboard. It handles the metadata of the analytics and data sources so that only necessary information is shown for the user. GraphQL is a middleware software framework from Facebook, which is used here to handle the communication between frontend components and analytics system. The main functionality of GraphQL is that it can buffer and bundle the calls from frontend to analytics layer. It also enables the stateless design for the UI so that waiting data for a single analytics component on the dashboard won't freeze the whole dashboard.

This deliverable is supported by deliverables D3.1 (Data requirements report) and D4.1 (Inventory of Stakeholder Required Data Analytics Methods). The work related to this deliverable is carried out in Task 5.1. The Policy Board of the project has a significant role for supporting the requirements specifications for this deliverable. The structure of the deliverable is as follows: At first, the identified stakeholders for each pilot are described and then the scenarios and user stories from the stakeholders are explained. After this, the available resources and data from each pilot to support the selected scenarios are described and from these the identified minimum requirements for the visualizations are introduced. The pilot descriptions are based on detailed descriptions from Pilot description Living document. At the end of this deliverable, the technical implementation plan of the system is defined for the first prototype. The deliverable D5.6 (Visual analytics tool(s) – MIDAS Dashboard UI evaluation 1) will report the user experience (UX) testing results of the first MIDAS Dashboard prototype.

2 Pilotwise KPIs

2.1 Basque Country

- The MIDAS platform should enable to understand and identify the etiology of the childhood obesity in the Basque Country and potential areas of intervention
- MIDAS platform should enable the Basque pilot to
  - To provide and clearly identify both crude and adjusted rates of target variables.
  - To provide information at different granularity levels, adequate for each stakeholder.
  - To cover time, place, and person analysis for epidemiological studies

2.2 Finland

Research question 1 (regional): How to use Midas platform and rich data from various sources in social and healthcare units in the regional stakeholder organizations to support preventive policy making?

- MIDAS should enable us to understand
  - what is the current practice for preventive policy making
  - to what extent data, digital services, analytics, visualizations are currently being used for aiding decision making?
  - what are the gaps in the policy makers decisions, how are the gaps affecting to their work?
  - What sort of data, analytics, visualizations could help decision makers in the preventive decision making?
  - Which data sources can be combined in order to create geographic and demographic analytics?

Research question 2 (national): What kind of regulations, enablers, metrics and should be enforced;

- How to identify correlations between the different parameters related to the policy maker decisions escalating on various government sectors;
- How can we simulate the different systemic correlations and impact factors to the policy makers to help their decisions?

2.3 Northern Ireland

- The MIDAS platform should enable analysis of the available datasets to identify effective preventative measures.
- The MIDAS platform should enable us to carry out a longitudinal analysis and track a cohort of Looked After Children as they move in and out of care, use a

variety of health services, and look at patterns of behaviours and changes over time. We are currently unable to do this analysis.

- The MIDAS platform should enable the use of visual analytics to assess the impact of changing variables and indicators on policy decisions
- The MIDAS platform should enable the harmonisation and integration of multiple data sources

2.4 Republic of Ireland

- KPI 1: The provision of an adoptable intelligent analytics platform for stakeholder use
- KPI 2: Identification of the cohort of persons with diabetes in the Republic of Ireland
- KPI 3: Mapping of the care journey for people with diabetes through the health service
- KPI 4: Determining of information that can be shared to facilitate better use of resources and services for people with diabetes
- KPI 5: Determining the best fit for diabetes services and their locations based on population based geographical needs
- KPI 6: Improving prescribing and reducing hospital admissions for people with diabetes
- KPI 7: Identification of improved outcomes if certain ‘data insight’ based policies are implemented
- KPI 8: Identification of an EcoSystem of policy based services for improved patient outcomes and the identification of partners to drive improved outcomes

3 Stakeholders and Pilot Descriptions

3.1 Basque Country

Childhood obesity has emerged as an important public health problem in Europe and other countries in the world, and according to the world health organization (WHO), the worldwide[1] prevalence of obesity nearly doubled between 1980 and 2008. In the WHO European Region 1 in 3 11-year-olds is overweight or obese (Health Behaviour In School-Aged Children (HBSC) Study)[2]. According to a childhood obesity review report[3], the increasing prevalence of childhood obesity is associated with the emergence of comorbidities previously considered to be "adult" diseases (e.g. type 2 diabetes mellitus, hypertension, or nonalcoholic fatty liver disease), tracks strongly into adulthood. It is the consequence of an interaction among a complex set of factors that are related to the environment, genetics, and ecological effects such as the family, community, and school. Basque Government, into its general lines of the XI legislature, within the health promotion and prevention action line pointed to the creation of specific child obesity prevention plan as one of the new two main action areas (details in Basque[4] or Spanish[5]). As such, childhood obesity prevention has been targeted as a prioritised regional policy for the Basque region within the MIDAS project. In the Basque Country, the identified key stakeholders for the pilot case are:

I. Minister of Health / Director of Public Health and Addictions,
II. Province-level public health epidemiology manager and,
III. Pediatrician responsible for an assigned children patient quota.

In addition to these, municipalities, academic institutions and other associations involved in programs for the prevention of childhood obesity will benefit from the new tools that will be developed to monitor the problem.

The Minister of Health / Public Health Director is the person responsible for defining and monitoring health policies, as well as presenting and reporting updates in health commissions and to the public in general. The province-level health epidemiology manager is responsible for monitoring health issues and the effect of implemented policies at province level, and for reporting to the public health director about the

1 http://www.euro.who.int/en/health-topics/noncommunicable-diseases/obesity/data-and-statistics

2 http://www.euro.who.int/en/health-topics/Life-stages/child-and-adolescent-health/health-behaviour-in-school-aged-children-hbsc/growing-up-unequal.-hbsc-2016-study-20132014-survey

3 http://www.mayoclinicproceedings.org/article/S0025-6196(16)30595-X/fulltext

4 http://www.euskadi.eus/eusko-jaurlaritza/-/albistea/2017/pertsonena-osakidetzako-langileena-eta-euskal-osasun-sistema-publikoa-unibertsal-zuzen-eta-kalitatezko-gisa-sendotzearena-izaten-jarraituko-du-xi-legealdiak/

5 http://www.euskadi.eus/gobierno-vasco/-/noticia/2017/la-xi-legislatura-seguira-siendo-la-de-las-personas-los-profesionales-de-osakidetza-y-la-de-la-consolidacion-del-sistema-publico-vasco-de-salud-como-universal-equitativo-y-de-calidad

province level health issues evolution and proposing candidate interventions (at province or country level) to be triggered through policies. Pediatricians responsible for an assigned children patient quota, are responsible for the health care of certain children. They are the direct contact for the child obesity policy being studied. They can monitor and be part of the intervention and also analyse their groups in comparison to other quotas. They can report on higher instances about possible interventions within their quota or a higher level. Table 1 summarizes the identified stakeholders for the Basque Pilot.

Table 1: Basque Pilot Stakeholders

Name	Role/ Affiliation	Patient quota level	Province level	National level
Jon Darpón	Basque Health Minister			x
Miren Dorronsoro	Director of Public Health and Addictions			x
Joseba Bidaurrazaga	Vice-Director of Public Health and Addictions of Biscay		x	
Javier Nuñez	Pediatric Endocrinologist		x	
Iratxe Salcedo	Pediatrician	x		
Pilar Aizpurua	Pediatrician	x		
Ainhoa Zabaleta	Pediatrician	x		
Jesús Arranz	Pediatrician	x		

3.2 Finland

The topic of the Finnish Pilot is “Preventive mental health and substance abuse of young people”. The topic was established after several interviews with policy makers in Oulu region. Mental Health America notes of studies that show half of those who will develop mental health issues show symptoms as early as by the age of 14[6]. The time between prenatal development and early adulthood is crucial for the brain and mental development. Policy improvements and resource investments in preventive services and early intervention programs can significantly impact school drop-out, homelessness, seclusion, addiction to harmful substances, unemployment scenario.[7] The World Health Organization defines mental health as “a state of well-being in which every individual realizes his or her own potential, can cope with the normal

6 http://www.mentalhealthamerica.net/issues/prevention-and-early-intervention-mental-health

7 http://www.mentalhealthamerica.net/issues/prevention-and-early-intervention-mental-health

Grant Agreement No: 727721

stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community"[8]. Therefore, in the Finnish MIDAS pilot context, preventive mental health refers to the approach that will incorporate preventive services and make it possible for early intervention to improve the general populations' mental health. Also, the World Health Organization defines substance abuse as "the harmful or hazardous use of psychoactive substances, including alcohol and illicit drugs".[9]

In the Finnish pilot, the focus is on young people, between the age of 7-24. This age categorisation is because young people show symptoms of mental issues from early childhood to their young adult age. Onwards in the Finnish pilot, the mental health state of young people was emphasized in most of the initial interviews and workshops with the local policy makers. The main reason for choosing this case is twofold. First, preventive mental health is a timely issue that needs intervention, often resulting in substance abuse, difficulties in finding education and employment. Second, the case provides a wide base for accumulating numerous variables and thus provides the scope for testing the MIDAS platform's analytical depth and dynamics in order to discover meaningful data on a systemic problem.

The Finnish Pilot considers the policymaking at national and regional level. At national level, decision makers include especially the Ministry of Social Affairs and Health domain staff, who are responsible for drafting the national policies of the future. At regional and city level, the Pilot concerns stakeholders and decision makers, such as directors of health and wellbeing services or social care services, but also directors of culture and education services. At both levels, these people are responsible for drafting recommendations and policy suggestions for politicians, who then process the final decisions based on the information provided from the experts of the targeted subject fields. Healthcare, social care and wellbeing are closely interlinked when referring to preventive care. Stakeholders for the "preventive mental health and substance abuse of young people" are identified to be located at three different layers of policy makers / facilitators: city level, regional level, and national level. Table 2 summarizes the identified stakeholders for the Finnish pilot.

Table 2: Finnish Pilot Stakeholders

Name	Role/ Affiliation	City level	Regional level	National level
Erkki Hämäläinen	Director of Education		x	
Leena Pimperi-Koivisto	Director of Joint Municipal		x	

8 http://www.who.int/features/factfiles/mental_health/en/

9 http://www.who.int/topics/substance_abuse/en/

	Authority			
Sirkku Pikkujämsä	Director of Healthcare	x	x	
Marja-Leena Meriläinen	Head of Health, Social and Education Services		x	
Elina Välikangas	Development and Quality Manager	x		
Heidi Alatalo	Finance Manager	x		
Kirsti Ylitalo-Katajisto	Director of Healthcare and Social Welfare	x		
Arja Heikkinen	Director of Social Welfare	x		
Liisa-Maria Voipio-Pulkki	Director, Ministry of Social Affairs and Health			x
Hannu Hämäläinen	Ministry of Social Affairs and Health			x
Taru Rastas	Ministry of Transport and Communications			x

3.3 Northern Ireland

The pilot topic for Northern Ireland is “Children in Care”. In this scope the term “Looked After Children” refers to a diverse group that varies in terms of age, ethnicity, the reason for being looked after, age of first entry into care and duration within care. However, it is fair to say that, while some children and young people in care can go on to enjoy success, as a group, educational and other outcomes tend to fall significantly below those of the general population. Such large shortfalls are not just concerning in themselves, but also as predictors of later life chances. As a group, looked after children are at far greater risk of experiencing social exclusion.

One of the most challenging aspects within the pilot topic is the ability to track children before they enter into care; the factors that contribute to them becoming looked after; movement to different placements whilst in care; referrals to other Health and Social Care (HSC) services; and their movement out of care. Currently the Department of Health (DoH) is restricted to six monthly and annual “snapshot” data (gathered at a single point in time) rather than longitudinal data. It is therefore difficult to establish patterns, historical context and plan the most effective services to support looked after children.

DoH has a statutory responsibility to promote an integrated system of HSC designed to secure improvement in:

- The physical and mental health of people in Northern Ireland;
- The prevention, diagnosis and treatment of illness; and
- The social wellbeing of the people in Northern Ireland.

DoH discharges these duties both by direct departmental action and through its 17 Arm's Length Bodies (ALBs). The Family & Children's Policy Directorate within DoH is responsible for the development of strategy, policy, and legislation in the areas of family support, children in need, child protection, looked after children, care leavers and domestic and international adoption. This project will benefit in particular from the policy relating to looked after children and the services put in place to minimise their entry into care in the first instance and those provided to them whilst in care.

The Department of Health (DoH) have responsibility for the development of policy regarding Looked After Children. They are supported by professional statisticians working in the Information Analysis Directorate (IAD). IAD produce a range of reports of publications using maps and graphs in static (PDF) format to monitor activity and performance. Implementation of policy is the responsibility of each of five Health and Social Care Trusts. Business Services Organisation (BSO) provide regional IT services to the Health & Social Care Trusts. This includes the management of the HSC Regional Data Warehouse which will be the main data source for this project. BSO will also host the NI implementation of the MIDAS platform. A full list of identified key stakeholders with their roles are presented in table 3.

Table 3: Northern Ireland Pilot Stakeholders

Name	**Role/ Affiliation**	**Policy development & oversight**	**Analysis to support policy dev & oversight**	**Data & IT provision**
Richard Pengelly	Permanent Secretary of DoH, DoH	x		
Eilís McDaniel	Director of Family & Children's Policy Directorate, DoH	x		
Eugene Mooney	Director of Information Analysis Directorate, IAD, DoH		x	
Malcolm Megaw	Head of Community Information Branch in IAD, DoH		x	

Heidi Rodgers	Deputy Principal Statistician, IAD, DoH		x	
Liam McIvor	Chief Executive, Business Services Organisation, BSO			x
Karen Bailey	Director, Customer Care & Performance and IT Services, BSO			x
David Bryce	Assistant Director IT Services, BSO			x

3.4 Republic of Ireland

The “Healthy Ireland” framework and related policies is a key delivery objective, and within the context of the framework the MIDAS focus is on the diabetes challenge. Currently a new diabetes dataset is being refined and scoped to identify and efficiently manage patients with diabetes across the general population in Ireland (Type 1 and 2). The aim is to have a single data platform for Analytical Insights and Data Visualisation of agreed datasets for diabetes with options for extending the platform to other long term conditions.

The key stakeholders are the Department of Health (DoH) for policy planning and many divisions of the Health Service Executive (HSE). The purpose is to plan a platform which supports policy decision making from a DoH level to public health and HSE Governance level, particularly in resource management of budgetary allocation and staff planning and allocation, analytical insights to support decision making and efficient healthcare delivery from a perspective of caseload managers.

Chronic disease management poses an issue for patients with one or multiple disorders. Diabetes is a fast-growing disease affecting 10% of the population of Ireland and has significant symptoms for patients and life changing factors. It is estimated that there are 190,000 people with diabetes in Ireland (Institute of Public Health in Ireland, 2007)[10]. Approximately 30,000 (15%) of these people do not have Type 2 diabetes, but either have Type 1 diabetes, or genetic or secondary causes of diabetes. The remaining 160,000 (85%) patients have Type 2 diabetes. A significant proportion of these patients (20-30%) remain undiagnosed. It is expected that the number of people with Type 2 diabetes will increase by 60% over the next 10-15 years. The challenge is to gain insights from existing data to make policy decisions

10 https://www.publichealth.ie/sites/default/files/documents/files/Making%20Diabetes%20Count%20What%20does%20the%20future%20hold.pdf

for the maximal impact on the wellbeing of these people and to drive operational efficiency.

This policy was selected as chronic disease management poses an issue for patients with one or multiple disorders. Diabetes is a fast growing disease affecting 10% of the population and has significant symptoms for patients and life changing factors. The MIDAS project proposes to afford meaningful information insights that can enhance policy. In essence, the success of MIDAS for the Republic of Ireland (ROI) stakeholder community will be determined by its ability to provide insights from existing data to make policy decisions for the maximal impact on the wellbeing of these people and also to drive operational efficiency.

It is necessary to identify all patients with diabetes and develop a population health plan for them supported by trend analysis from data patterns would be crucial. This will improve financial costings, and enable moving from treatment costs to prevention costs and increase health factor for patients. It is also necessary to identify the effectiveness of diabetic medication and how hospital admissions can be avoided by effective prescribing. Demographic insights, age, gender and socio-economic dimensions are also central to the analytical process. Information collated will be utilised to determine additional resources required to drive improvements in outputs, including staffing, care pathways and intervention frequency.

There is no single centralised diabetes register for Ireland. Sharing and merging of appropriate data sets from various sources is an issue. This involves merging the datasets on a single dashboard to provide a picture of diabetes in Ireland. Completing the picture with accurate and up to date data for policy and decision making is the goal.

Within the Republic of Ireland, the HSE stakeholders are Pat Kelly, Peter Connolly and Yvonne Goff who are leading the policy aspects initiative. Within the Department of Health, Muiris O'Connor, who is the Head of Research and Development and Health Analytics, is keenly interested in the value of the initiative to drive policy from evidenced based data. Data Management is being led by the corporate delivery of programmes, enterprise architecture, and the Chief Clinical Information Officer (CCIO). A clinical focus is offered from a clinical endocrinologist and alignment with the National Clinical Programme for Diabetes. A summary of identified key stakeholders is presented in table 4.

Table 4: Irish Pilot Stakeholders

Name	Role/ Affiliation	Local Level	Regional Level	National Level

Muiris O'Connor	Assistant Secretary DOH			x
Pat Kelly	HSE, Corporate Delivery Director, OoCIO			x
Peter Connolly	HSE, Head of Enterprise Architecture & Design Authority			x
Yvonne Goff	Chief Clinical Information Officer (CCIO), HSE			x
Dr Paul Davis	Head of Management Group, DCU			x
Prof Anthony Staines	Clinician and Epidemiologist, DCU			x
Regina Connolly	Digital Health Innovation and Information Systems Governance, DCU			x
Paula Kavanagh	CCIO Coordinator and Diabetes Lead, HSE			x

4 Scenarios and User Stories

The scenarios and user stories varies significantly between different pilots. This is considered as an advantage as it provides a rich field for testing the MIDAS solution with heterogeneous user groups and use cases. On the other hand it also raises challenge for the system to support them all.

4.1 Basque Country Pilot

The actual scenarios and user stories are being researched and developed within MIDAS in the context of a specific Health Plan, which is on developed to tackle children obesity. There are also other related government initiatives as recently introduced “Healthy Nutrition Plan”. The Health Promotion, Health Surveillance, Health Protection policy processes are involved in the following scenarios.

4.1.1 Descriptive or Effectiveness assessment scenario

Nowadays, the reality of children overweight and obesity is captured through surveys covering a limited sample and a limited time granularity. Health managers lack of up-to-date detailed descriptives of the population. Moreover, they cannot do a reliable analysis of policy implementation pilots effectiveness.

In the Basque autonomous country, the most comprehensive obesity data about children's obesity available nowadays comes from the Health Survey. This survey is made every 5 years and it consists in a large questionnaire about many health topics, completed by a sample of the population. The results of the survey are presented as written reports, where you can find static table and graphs. The last reports where published in 2015 and they came from the survey of 2013. Other specific studies are also done but are limited in specific areas and / or ages.

4.1.1.1 User Story

The main user story within descriptive or effectiveness assessment scenario is, that health managers at different levels can have an updated descriptions of their aggregation level and contrast it with others at their aggregation level or ones at higher levels of aggregation. These decision maker levels can be for example:

- Individual,
- Patients-group assigned Pediatrician (Patient quote),
- Province-level public health epidemiologist and
- Country-level public health manager.

Additionally, visualising the evolution of these descriptive visualizations with implemented pilot policies would help for further considerations with them.

4.1.2 (Inter) variable correlation or relationship identification scenario

Currently, it is not clear which factors are responsible or affect in the process of a children evolving to an overweight or an obese state. Therefore, intervention areas are not clear and interventions are non-data (or limited-data) grounded. The interventions usually target intervention areas that are:

- established within general practice;
- reported in the scientific state of art; or
- hypothesised by experts' experience and knowledge.

4.1.2.1 User story

To be able to answer the scenario it should be possible to identify

- the relationships (one to one and more complex ones) among predictor variables (i.e. health and non-health variables), and
- the variables representing the children overweight and obesity.

These identified relationships (e.g. through correlation analysis, or regression) should be visualised for domain experts, so they can concentrate on some interventions areas. In addition, domain experts should be able to choose for visualization those that make sense (avoiding non-spurious correlations) and are targetable by the Public Health manager. An example visualisation can be a correlation matrix. Additionally, it would be desirable to identify clusters of children that have similar profile and patterns that lead to similar values of variables representing children overweight and obesity.

4.1.3 Forecast or Monitor policy scenario

An evolution of the available tools for the properties introduced in the descriptive user case story, will allow responding to question "What's the current status?" The next evolution of the tools for the variable correlation/relationship identification, will allow responding to the question "Which should be my policy's intervention areas?" These do not allow for predicting the evolution of children's overweight and obesity health issue, nor analysing data model-based policy resulting scenarios.

4.1.3.1 User story

This user story will allow health managers to forecast short-medium-long term evolution of children's overweight and obesity health issue based on historical data and living updated data. These actions target also to data model-based forecasting of scenarios, where intervention policies can be chosen based on improvements on policy-expected predictor variables, which would lead to optimum future scenarios. Additionally, monitoring of the evolution of an implemented data model-based policy could be targeted for feedback.

4.2 Finnish Pilot

In Finland, the pilot is concentrated to two main scenarios and user stories within them. The regional scenario focuses on local health decision making and national scenario focuses on general health regulations, which also affect to the regional case. On timeline, the regional scenarios focuses more on short-term action when the national focus is on longer-term effects.

4.2.1 Regional scenario

The regional case has a more short-term and operative focus, though long-term aspects are included as well. The main research question is:

"How social and healthcare units in the regional stakeholder organizations should conduct personnel and financial resource allocations for preventive services based on rich data from various sources?"

The regional scenario includes stakeholders from city and regional levels; meaning: the stakeholders from City of Oulu, Kuusamo, Kempele, Oulunkaari and their decision making situations are more directly considered in this scenario. While the regional scenario focuses on the personnel and financial resource allocation decisions for preventive services, currently this decisions are made based on national indicators, periodic information sourced from partners/ paid vendors (e.g. monthly excel based reports from addiction care services), primary data collected from meetings and customer orientation (mostly qualitative), and OukaDW for some specific cases. The Social care and Health and welfare services of the City of Oulu use also a high level follow up reports a periodically (every 6 months). The challenge is that regional decision makers face the obstacle arising from confidentiality and data protection. While the amount of data is yet not so significant in most of the stakeholder site, individuals are possible to be identified which creates the conflict against confidentiality and data protection.

4.2.1.1 User story

In regional level, the main decision making scenario involves the allocation of financial and personnel resources and services in city sub-regions based on the trends and indicators that emerge from the field. The decision-making is supported by the data collected from the state of healthcare services and people's wellbeing in different sub-regions of the city. Supporting personnel synthesize this information, and provide reports for the city council and board as well as for the regional administration who make decisions on city and regional healthcare services level. The stakeholders of the city need analytics and visualization tools to show for the politicians the short term and long term effect of the applied policies and to illustrate the impact of policy decisions on the wellbeing of citizens. From a service-organization and distribution related decision-perspective, budgetary and service performance related analyses form an important aspect for efficient resource utilization.

4.2.2 National scenario

The focus of the national case is long-term and strategic. On the national level, the Finnish MIDAS pilot intends to introduce data-driven foresight, which eventually facilitates preventive policy making. The national level scenario is quite relevant to the regional level scenario; because, while the regional level policy makers focus more on specific policies (resource allocation, special campaign, etc.), the national level policy makers focus on creating legislations enabling factors that affect regional policies as well, in addition to some international perspective. The main research question is:

“What kind of regulations, enablers and metrics should be enforced and how in order to achieve the systemic targets escalating on various government sectors?”

Situations visualised by system dynamic models in the national case include:

- Evaluating the impact and side effects of more widely used personal data in preventive actions
- Visualizing the ripple effect of partial optimization on one government sector to other sectors

Currently, in Finland, the healthcare sector is in the middle of a huge change due to the social and healthcare reform. It is going to change the responsibility of organizing services and financial resources of healthcare from 326 municipalities and 20 hospital districts to 18 counties. Currently each municipality and hospital district have their own decision support systems and the way these systems are built varies a lot. There is also an ongoing development to modify legislation related to the use of different data sources on national level. Based on the interviews, we understand that the new legislation is constructed in a way, which is compatible to EU data protection regulation (GDPR).

The availability of data is not the challenge on national level in Finland, but refining it for proper application is. Since the early 90’s in Finland, municipalities have been responsible for all of it, including the most demanding specialist health care. According to the Finnish constitution, municipalities have the general authority on things that they are legally required to provide. This means that on the state level, there has been no need and in fact not even a mandate, for more detailed data-based governing.

4.2.2.1 User story

On national level, decision-making involves individuals involved with political parties and apolitical civil servants supporting them. Decision-making foregrounds differ sometimes within individuals with varying professional goals. Decision-making strategies can also vary for a politician over time, depending on personal and political goals in respect to other parties. Typically, materials developed and brought by apolitical civil servants support legislative decisions. In this context, strategic foresight through system dynamics modeling and simulation will prove to be a handy tool to create legislation that does not only protect political goals, but also advances preventive decision making in regional levels. In addition to strategic foresight, it has been acknowledged that more suitable analytics and visualization techniques can foster policy formulation in national level creating a coherent overall scenario by connecting it to the regional cases.

4.3 Northern Ireland

4.3.1 Scenario

The term “Looked After Children” refers to a diverse group that varies in terms of age, ethnicity, the reason for being looked after, age of first entry into care and duration within care. However, it is fair to say that, while some children and young people in care can go on to enjoy success, as a group, educational and other outcomes tend to fall significantly below those of the general population. Such large shortfalls are not just concerning in themselves, but also as predictors of later life chances. As a group, looked after children are at far greater risk of experiencing social exclusion.

The DoH work is challenged by continued downward pressure on the HSC budget and introduction of Welfare Bill, which will have an impact in family incomes and may adversely affect those currently on “the edge of care”. Higher level of reported incidents of domestic violence has been attributed to these changes. Handling this challenge requires new tools and methods. The Health and Social Care Board (HSCB) is responsible for operational policy and processes arising from the regional strategies. Each of the HSC Trusts will then develop their own local policy and procedures.

The current relevant strategy and policy for the Pilot topic are described in two documents:

- “Care Matters in Northern Ireland”, which outlines a strategic vision for wide ranging improvements in services to children and young people in and on the edge of care; and
- Families Matter: Supporting Families in Northern Ireland Regional Family and Parenting Strategy

Where new strategies are in development for topics:
- Looked After Children
- Parenting and Family Support

The Legislative framework connected to the topic includes regional and international laws and regulations. The current regional legislation:
- The Children (Northern Ireland) Order 1995,
- Children (Leaving Care) Act (Northern Ireland) 2002,
- The Children (Leaving Care) Regulations (Northern Ireland) 2005, and
- The Adoption (Northern Ireland) Order 1987

The international regulations which must be taken account are:
- UN Children Rights Convention and
- Hague conventions 1980 and 1996

The key outcomes of the scenario would include:
- Being able to track patterns of behaviour over time;
- Creating visualisations of data about movement of children in and out of different types of care;
- Being able to track Adverse Childhood Experiences into the future helping to identify how we address their needs (and whether this was effective); and
- Examine trends in data to help HSC professionals to enquire more deeply (policy informing practice).

4.3.2 User stories

4.3.2.1 Policy development & oversight

The Department of Health (DoH) have responsibility for the development of policy regarding Looked After Children. They are supported by professional statisticians working in the Information Analysis Directorate (IAD) as the staff in the Policy Directorate do not perform any data analysis by themselves. The actual implementation of the policy is the responsibility of each of five Health and Social Care Trusts.

An intuitive visualisation platform from MIDAS has the potential to be used by:
- The Director of Family and Children's Policy;
- Policy leads who are responsible for the development of policy and strategy relating to family support, children in care, adoption, safeguarding, child protection and care leavers;
- social services officers who provide advice on professional social care matters;
- HSC service commissioners and practitioners;

Actions emanating from the draft Looked After Children Strategy will be prioritised. In particular, the ongoing Review of Secure Care and its interface with other regional specialist children's services, will enable the review team to track the cohort of children who frequently are referred to these services and help to identify patterns of behaviour. This will help to inform and consolidate the Review's findings and in the longer term, allow for the exploration of variation in HSC practices with regard to this group of children and to identify lessons to help inform policy development and service design.

4.3.2.2 Analysis to support policy development & oversight

IAD produce a range of reports of publications using maps and graphs in static (pdf) format to monitor activity and performance. Currently, the used tools include SPSS, SAS, Excel, MapInfo and SAP Business Objects. The current mandatory requirements for policy oversight and development are the reports on Delegated Statutory Functions, which are formally submitted twice yearly. The developments of policies are based on these reports.

Much of the data is extracted/exported from the Regional Data Warehouse by running SAP Business Objects reports. Due to lack of training/skills the statisticians in IAD do not use the capabilities of Business Objects for analysis. There are also some manual (spreadsheet) returns from Trusts. Other governmental or Open Data sources such as NI Multiple Deprivation measures and the Central Postcode Directory are also used. The datasets are typically uploaded and analysed by using Stata, SPSS and/or Excel. Static pdf reports are produced for distribution to the Policy Directorate.

One of the most challenging aspects is the ability to track a child before their entry into care, the factors that contribute to them becoming looked after, movement to different placements whilst in care, referrals to other HSC services and their movement out of care. DoH is restricted to 6 monthly and annual "snapshot" data (gathered at a single point in time) rather than longitudinal data. It is therefore difficult to establish patterns, historical context and plan the most effective services to support looked after children.

With MIDAS platform the data should be accessed with direct links to data sources which will remove the need for time consuming manually instigated exports and uploads. The main users for new visualization and analytics system would be the statisticians and public health officials.

4.4 Republic of Ireland

Within the Republic of Ireland, to drive policy, evidenced based data is required. The ever-increasing population of persons with diabetes is a challenge from a treatment and preventive perspective. Presently, a true picture of the number of persons with diabetes cannot be identified, and in planning the future how to best allocate resources for the purposes of best possible health outcomes and ensuring efficiency. For policy makers to reflect on the resources that should be allocated to both treatment and prevention, it is necessary to analyse numbers and locations of where the disease is manifesting. A focus on extraction of data from a range of sources is a priority to identify all persons with diabetes who may require healthcare interventions. At present data can be obtained from silo data collection mechanisms (HIPE Discharge data, PCRS prescribing data) and from coding in GP Systems around the country. The whole picture is required to identify the true total of persons with diabetes to track and plan care and treatment.

The challenge is to regularise a national diabetes register with factual information. Having a unique identifier to validate each person and to match them with other data sets and information is critical. Data sharing and ownership of the register would need to be governed to ensure data remains current for policy makers to react and change plans accordingly.

4.4.1 Forecast or Monitor policy scenario

Forecast Tools for the Republic of Ireland scenarios are critical to shape policy within the Dept of Health and at policy level within the HSE. We wish to facilitate forecasting of a national picture of diabetes across Ireland, geographical area, urban and rural and by other correlation trends such as deprivation patterns and medication trends. Predictive modelling to support policy makers to plan diabetes programmes based on preventative measures rather than treatment options is a core component. Caseload management of persons with diabetes from an efficient staffing perspective as well as healthy lifestyle for individuals is key.

4.4.2 Scenario

The Republic of Ireland (ROI) scenario is based on access for policy makers to a national diabetes register with up to date information on all persons with diabetes in their region.

The aim is to support self care for patients with diabetes through clinical health care professionals sharing information relating to diabetes care in real time providing supported decision making for the person with diabetes.

Higher level policy makers need information from a range of sources amalgamated in a single workspace or dashboard to support decision making and require quality information which is evidenced based and supported by quality standards which gives valuable information to determine the level of risk in a community, in the hospital on any given day or period.

Issues in the ROI are the lack of:

- Access to health records at the point of care or lack of sharing of information across hospital/community boundary
- IT software to facilitate information recording in a health record which is accessible for others
- Database to support caseload management (prioritisation by risk and health needs)
- Statistical and trend analysis of care episodes to determine risks
- Access to standardised quality data, following set criteria for collection and validation
- Real time data which can be comparatively analysed and utilised and trends identified

4.4.3 User Stories

4.4.3.1 Regional Story

Currently diabetes registers exist in single GP Practices and in acute hospitals where persons with diabetes undertake treatment. This is not a complete picture and is not visible in a single register in order to uniquely identify the number of persons with diabetes in Ireland (Type 1 and 2) and their treatment and care plan. Visualisation and analytics of the data with trend analysis would offer core clinical information on the validity of treatments and education plans. It would also assist in caseload management to place resources such as clinical nurse specialists and endocrinologists in areas which have a high density of persons with diabetes.

4.4.3.2 National Story

Nationally there are significant data available from national data sets such as HIPE and PCRS but in isolation, this does not give a true picture. With the onset of telehealth, there are opportunities to gather data from glucose monitoring machines and fitness data relating to their health from wearables. Acute hospitals also store information on patients with diabetes registered in their region on Electronic Patient Record (EPR) such as ProWellness. Eye and foot screening programmes are coordinated through the national clinical Programme and Working group which could also offer valuable data sets. Once individual data sets are identified, data modelling can be activated to provide for resource planning and caseload management for persons with diabetes. At a national level, financial resources can be mapped according to population trends and incidences and at a regional level, decision support can be offered to healthcare professionals to plan health wellness of targeted populations.

5 Resources and Data variables

This section describes the resources including the possible data sources in different pilots. Here the descriptions focuses on visualization methods and visualization requirements of stakeholders. Detailed descriptions are documented on Pilot description living document. Details and descriptions of data sources are reported in deliverable D3.1 Data requirements report 1. MIDAS Dashboard does not handle the integrated health data or analytics on top of it , MIDAS Dashboard only visualises the results from the analytics layer, which gets the integrated data from data integration layer of MIDAS Platform. The data sources for the integration layer can be the actual healthcare systems or flat files. Each Pilot will have individual deployment of the system. In what regards the MIDAS-customized access to open data, the acknowledgement of its impact is transversal to all the stakeholders, and efforts are in place to better fit to the needs of each pilot, interacting the extracted information with the data available at their site. Thus, a development into a second version of the pilot is needed to see progress in the integration of open data benefits.

5.1 Basque Country

5.1.1 Infrastructure

The anonymised and secure data sets provided to MIDAS pilot will be located in a private cloud provider by Euskaltel telecom operator. The whole MIDAS data analytics system deployment for the Basque Pilot will be located on the same environment. BIOEF will be the data and system controller in the Basque Pilot of MIDAS. Access to the platform will only be granted to MIDAS partners who have signed a data access agreement. Data access agreement and procedures will be defined to fulfill the GDPR directive and Spanish regulation compliance, contrasting the approach with the regional data protection agency and ethical committee.

5.1.2 Human

The Basque MIDAS Pilot is led by Basque Foundation for Health Research and Innovation (BIOEF). BIOEF is the intermediary and provides access to Basque Government Public Health and to Osakidetza public healthcare provider. VICOM is actively collaborating with the BIOEF regarding data variable selection, user requirement identification and the general pilot description, contributing by including the technical perspective in pilot meetings.

As well as BIOEF's intermediary and access providing role, BIOEF is leading the ethical and privacy requirements understanding and fulfillment, as well as the Basque policy pilot deployment. Additionally, BIOEF is making contacts with third parties (e.g. a grocery chain and telecom operator) to identify and enable the use of complementary data sources in subsequent iterations of MIDAS.

The active involvement of people from Basque Government Public Health ensures the user requirements are well understood and results are oriented to meet their expectations.

Osakidetza's contribution is twofold. On the one hand the involvement of their corporate IT delivery people guarantees the correct health variable selection and understanding of selected data. On the other hand, Osakidetza allows us to engage with people involved in providing healthcare (A Basque pilot stakeholder).

5.1.3 Data

The Basque pilot includes two main data sources for the different analytics (descriptive visualization, variable relationship and forecasting) which are the public health provider's health data export (including individual's (and where available mother) characteristics, diagnosis, visits, measurements and prescriptions) and government open data (i.e. census data, environmental quality data, city planning). Other data sources are being considered for inclusion, and are being analysed and negotiated with third party data sources (i.e. grocery buying and geo-spatial movement-patterns and the geospatial data).

5.2 Finland

5.2.1 Infrastructure

The anonymized and secured data sets provided for the for the Finnish Pilot will be located on a virtual machine running in the safe ePouta IaaS service of the IT Center for Science[11]. The virtual server system will be running within Oulu University network and the whole MIDAS data analytics system deployment for the Finnish Pilot will be located on the same environment. University of Oulu will be the data and system controller in the Finnish Pilot for MIDAS. Each data provider will accredit University of Oulu to maintain the data sets in the MIDAS project. The ePouta environment, where the data is uploaded, has limited access only from the University of Oulu network by selected personnel and named personnel of MIDAS partners. The final use of the data is protected by a separate data transfer agreement required from all MIDAS partners, which limits the use of data only to MIDAS studies and forbids researchers to identify individual persons/people from the data. Only persons whose name is listed in agreements are allowed to access the MIDAS data in the dedicated ePouta location.

5.2.2 Human

From a human resource perspective, the Finnish MIDAS Pilot is organized with personnel from University of Oulu and VTT Technical research center of Finland. In

11 https://research.csc.fi/epouta

addition, personnel from THL are involved in the pilot as a Policy board member. Human resources from University of Oulu initially focused on the non-technical side of defining the nature of the pilot, creating the two pilot scenarios (national and regional), gathering user requirements and understanding current practices of policy making and how data is currently being used. Also, University of Oulu resources play a key role in studying the MyData approach potential within MIDAS as a whole. On the other hand, VTT researchers are tackling MIDAS activities from a more straightforward technical perspective. First, one group of researchers at VTT are responsible with the MIDAS visualization platform. Second, another group is participating in the national level pilot scenario with strategic foresight, system dynamics modeling and simulation. University of Oulu researchers are participating also to the national and local level pilot scenarios creation, and collaborating with local policy makers in the city of Oulu and in the Oulu region.

From a data related perspective, University of Oulu's MIDAS personnel are initially in contact with Northern Finland Birth Cohort (NFBC) research for the data, as the NFBC data is owned by the University of Oulu. Similarly, they are also collaborating with the city of Oulu for the data. VTT will work closely with THL to create the secured and anonymized dataset for MIDAS project. The developed data set will be owned by THL and operated by the University of Oulu in the MIDAS project according to a separate data transfer agreement between the University of Oulu and THL. VTT researchers are consulting the data definitions within all datasets. University of Oulu is also the data holding partner in the Finnish Pilot and will define the data access agreement for the Finnish Pilot.

5.2.3 Data

The Finnish pilot includes three main data sources which all have different role in analytics. These three sources are Northern Finland Cohort 1986 maintained by University of Oulu, the regional (city level) data from the city of Oulu data warehouse and national health register data from THL.

The Northern Finland Birth Cohort (NFBC1986) was begun in 1985 and was originally called as "The mother-child cohort study of morbidity and mortality during childhood with the special purpose of preventing mental and physical handicap". MIDAS data set from the NFBC1986 data includes selected variables from the whole data over the years 1985 - 2016. Some cohort data variables of the cohort are to be evaluated within MyData potential. The persons in birth cohort are currently 31 years old and the history of the cohort persons forms a solid base to develop and test the potential of MyDATA approach in MIDAS. The idea is to use cohort data to simulate personal data as the consent of the individual is secured for research.

The MIDAS data set from City of Oulu is collected from selected registers and data sources of the data lake of the city, called OukaDW (data warehouse). Data is selected from population, which has been 7-24 years old at year 2000. The selected data includes data from years 2000 to 2017 from these patients or when available later on during MIDAS time (possible updates). The variables are aggregated on postal code accuracy. Data from city of Oulu gives a detailed overview of how the policy making takes effect on regional level in local level.

The THL data set is collected from selected THL registers. Data is selected from the population, which have been 7-24 years old in the year 2000. The data is selected from these persons over the years 2000 to 2016. The data is selected from the population of the whole country and is anonymized before being transferred to the MIDAS environment. THL data enables MIDAS to compare and present the differences and changes in the policy target area in between different regions around Finland. The THL data set for MIDAS is owned by THL and will be operated by University of Oulu in the MIDAS project. After the MIDAS project ends, THL will study options to secure the dataset so that it can be released as open data in SotkaNET.

In addition to what has been described previously, the Finnish pilot will study possibilities to use open data from the SOTKANet maintained by THL. Under special interest are the KUVA indicators which include indicators computed by the Finnish Ministry of Social Affairs and Health from the data of THL, Statistics Finland and other environmental and health data. Detailed description of the data sets are found from D3.1.

Additionally, the Finnish pilot will run social media campaigns using the Twitter platform at first in collaboration with IBM. Through this user story, we will test the usability of social media data in health policy making.

5.3 Northern Ireland

5.3.1 Infrastructure

BSO will host the NI MIDAS platform in the HSC regional data centres. This will include the provision of virtualized servers, storage, resilience and back-up arrangements and connection to the HSC network. BSO shall control access to the server environment. The HSC ICT Security policy and Statement of Compliance[12] shall apply to all users.

12 https://www.health-ni.gov.uk/publications/hsc-ict-security-policy-version-119

Access by other MIDAS partners will be subject to agreed terms and conditions and on a named user basis. Remote access by such named users will be possible.

5.3.2 Human

Department of Health (DoH) policy leads will be key to agreeing the user stories and visualisation methods.

DoH Information staff will use their expert statistical and business knowledge and skills to assist in the development and quality assurance of the MIDAS platform.

BSO provide regional IT services to the NI Health & Social Care Trusts. This includes the management of the HSC Regional Data Warehouse, which will be the main data source also for this project. BSO will be responsible for supplying the dataset and hosting the MIDAS platform. BSO is in the process of appointing a temporary Project Manager to coordinate BSO activities and act a point of contact for the project in NI. Staff from various teams in BSO’s IT Services will be involved in the hosting services and in the data extraction. Technical partners in MIDAS are supporting BSO with system deployment and brining the scenarios and user stories in to action with MIDAS platform.

5.3.3 Data

The Health and Social Care data will be supplied to DoH under the HSC Honest Broker Service arrangements. The dataset will be pseudonymised by BSO prior to being made available to the MIDAS platform. Access to other HSC users outside DoH may be subject to an individual Honest Broker Service application which will be handled by the BSO.

No personal identifiable data will be available to the end users of the platform.

The Health and Social Care data is updated on a daily basis in the Regional Data Warehouse using a complete drop and rebuild approach. The data in the MIDAS platform needs to be updated preferably daily but at least on a weekly basis. The Health and Social care data relates to around 800,000 individuals. The used dataset will have a size around 10GB.

A detailed description of the Health and Social care data is found from D3.1. This may be supplemented by other governmental data sources such as Multiple Deprivation Measures.

5.4 Republic of Ireland

5.4.1 Infrastructure

Data will be anonymised and made available with secured deployment in Microsoft Azure where also the MIDAS deployment of pilot will be deployed. This will only be available to the relevant MIDAS participants who have signed the HSE's Service Provider Confidentiality Agreement and returned the relevant data processor material.

5.4.2 Human

No full time staff will be made available to this project from HSE. Resourcing is on a good will and best effort basis. If early insights from the data show merit in adding additional committed people resources to this project then this will be addressed. Technical partners in MIDAS are supporting HSE with system deployment and brinning the scenarios and user stories into action with MIDAS platform.

5.4.3 Data

HIPE Data records discharges from acute publicly funded hospitals in Ireland. It is a summary of the discharge and captures administrative, demographic and clinical information on the episode of care. (Clinical data =principal diagnosis classified using ICD10AM8th Edition and principal Procedure classified using ACHI).

Patient data with Diabetes Mellitus is captured using E10 Type 1, E11 Type 2, E13 Other, E14 unspecified, E09 intermittent Hyperglycemia.

PCRS Prescribing data is also made available for overlay with the HIPE datasets. This includes anonymised population datasets, drug products, listing of Hospitals, Community Health organisations and Pharmacies in Ireland.

A detailed description of the data is found in D3.1.

5.5 External Resources

5.5.1 Social Media Analytics

IBM is currently using Twitter to provide insights about the public's sentiment towards health policies. In future iterations we will look to include other sources, for example, Facebook, Linkedin, Google+, Instagram, Reddit.

IBM's Twitter chatbot will ask members of the public a series of questions about the given health policy. The entire response object from Twitter will be stored and augmented further with annotations from IBM's Watson services on Bluemix. An example of the Twitter direct message object can be found from their webpage[13]. Watson will then add sentiment and emotion values to the object, perform named entity extraction, concept extraction, identify people names and identify place names (possibly more in future iterations).

5.5.2 MEDLINE Analytics

The freely available medical/scientific research dataset MEDLINE[14] comprehends a large cover of that worldwide research (over 26 million citations) and is recognized as an important source of information in the daily life of both Public Health and Healthcare professionals. In that, MIDAS partner QUIN is developing an interactive text-mining tool generating several visualisation modules enabling the user to extract meaningful information from MEDLINE. Currently, QUIN is focusing on specific tools based on the existing state-of-the-art technology refocused to extract meaningful information from MEDLINE using the underlying MeSH (Medical Subject Headings) ontology-like structure. The analysis of the MEDLINE dataset is performed through WP3 and WP4 tasks. MEDLINE data is indexed with ElasticSearch and made available to analytics and visualisation tools.

5.5.3 Newsfeed Analytics

QUIN is developing several Public Health focused tools based on state-of-the-art text mining technology. Those are fed by the open data sources of (i) digital multilingual worldwide news, and (ii) MEDLINE (relating to 4.5.2). The common challenge of using free text documents to support public health decision making through visualisation tools is to analyse those texts and provide it with relevant metadata. The metadata attached to text documents describes the document content in a uniform and standardised way. For the above mentioned sources such metadata consists of geographical locations, types of events, health-related categories, source identification and timestamps. As described in the deliverable D3.1, Section 3.2, QUIN provides access to newsfeed.ijs.si through an agreement with Jozef Stefan Institute (JSI), Ljubljana. Newsfeed provides a real time stream of annotated, multilingual news. This stream is being further annotated with MeSH with modules developed in WP3 and WP4.

13 https://developer.twitter.com/en/docs/direct-messages/sending-and-receiving/api-reference/get-messages

14 https://www.nlm.nih.gov/pubs/factsheets/medline.html

6 Visualization Methods

The MIDAS Dashboard is the main interface for end users using the MIDAS Platform. It connects the analytics applied on local data and the results from the analytics applied for external resources. It is linked to management dashboards for the external resources and provides easy to use experience for all user groups of end users.

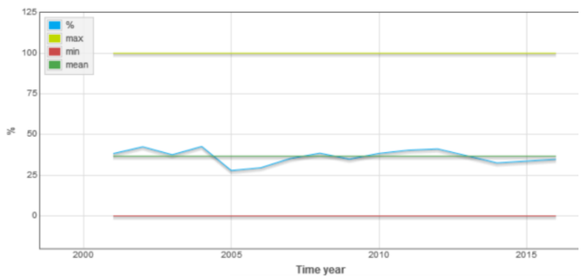
The main differences between pilot regions are that each pilot of the MIDAS Platform includes different data and some dedicated analytics and possible dedicated visualizations for them. Visualisation methods are applied on the MIDAS Dashboard via plugins, for which reason no modifications are needed to the actual framework between pilots. Only the used plugins differ between pilots. Common to all pilots are the basic analytics and visualisations. This chapter describes the identified common visualization methods as well as dedicated visualizations separately from all Pilots. In addition, the visualization schemes of external resources are described.

6.1 General Methods

This section describes the initial common visualization requirements over all Pilots. Those are tightly connected to the data analytics requirement behind the visualization. General analytics methods are described in deliverable *D4.1 Inventory of Stakeholder Required Data Analytics Methods.*

Through initial research over the pilot topics, personas, scenarios and wireframe examples, three main exploratory data analysis categories were selected for the implementation of the first prototype of the MIDAS platform. These categories were descriptive analysis, time series analysis and regression methods. Based on these, the visualization types in Table 5 were identified to get requested from each pilot. These were selected as initial visualisation types to be implemented in the first prototype phase for all Pilots in the MIDAS Dashboard:

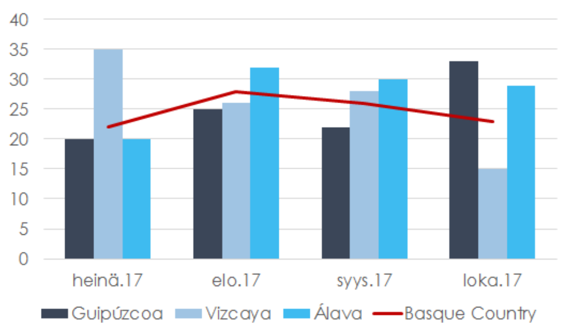
Table 5: MIDAS Prototype V1 Visualizations

Line graphs[15] display quantitative values over certain continuous interval or time period. The graph is typically used to analyze trends and how values have changed over time.	

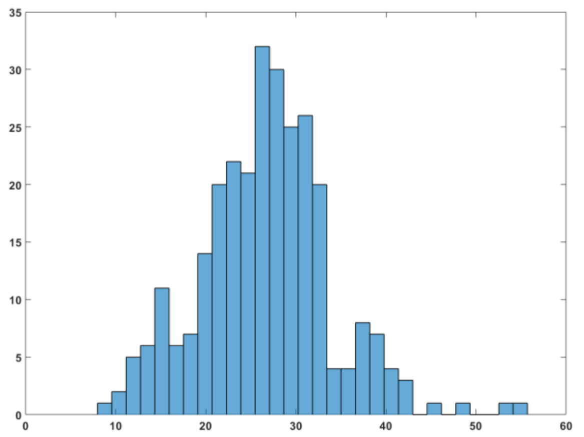
15 https://datavizcatalogue.com/methods/line_graph.html

Grant Agreement No: 727721

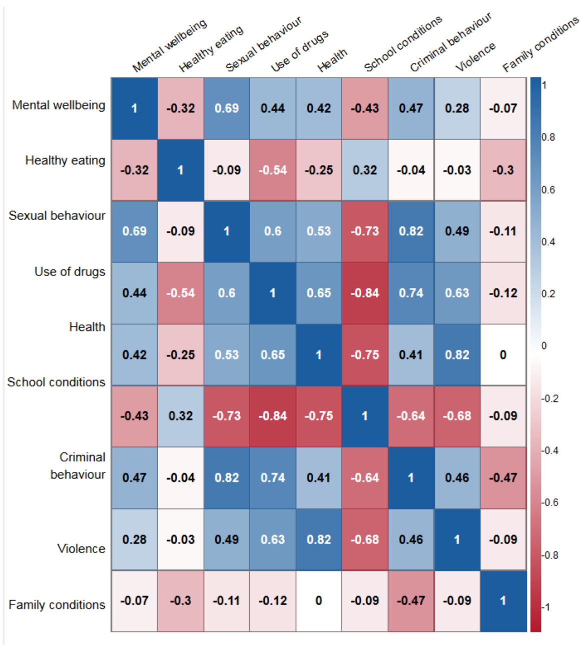
Bar Chart[16] can be drawn with horizontal or vertical bars. Sometimes horizontal chart is called as column chart. It is used to show discrete, numerical comparisons between categories. One axis defines the specific categories and another represents a discrete values.



Histogram[17] can be treated as specific case of Bars Charts where there are no actual categories but the values show the frequency of events in the studied interval or time period. Bar charts answer typically to question: "how many?", when histograms can show where the data is concentrated (in time or frequency).



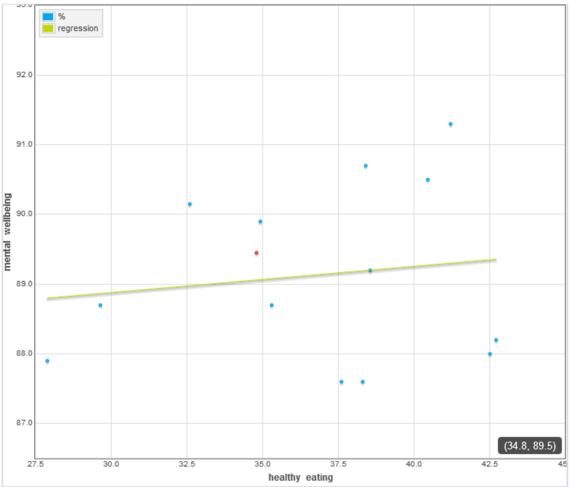
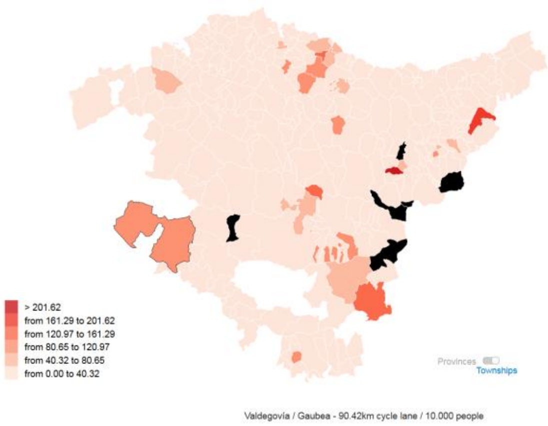
Heatmap matrix[18] present differences of the matrix values with color coded table. They are typically used with correlation matrices to study connections between variables. Common way of drawing heat maps defines one category of variables on rows and another category of variables on columns. The value matching on table is the value computed from these variables. Typical mathematics with heat maps are correlation or indicator type of analytics. The used colors on the table depends on the used color scale. For this reason a legend is often required with Heatmap in order to understand the content.



16 https://datavizcatalogue.com/methods/bar_chart.html

17 https://datavizcatalogue.com/methods/histogram.html

18 https://datavizcatalogue.com/methods/heatmap.html

Scatterplots[19] are drawn by placing points in a x-y-canvas. Typically, x- and y-values of points are two variables which possible relationship is studied. Often a line graph is drawn on top of scatter plot to visualize regression or trend line computed from the values of x- and y-axes. Scatter plot can be used to detect the strength or type of relationship between investigated values. Sometimes it is also used to detect outliers from the data.	
Choropleth Maps[20] include divided geographical areas or regions on a map, which are coloured, shaded or patterned to represent some value computed from the data connected to the matching region. It provides method to visualise values over a geographical area, which can be used to detect for example variation or patterns across geographical location.	

Based on the communication with end users, these figure types can cover most of the typical needs on data and analytics visualisation needs for policy makers within all pilots. These can be reflected to the following section 6.2, where the current visualisation experiences of the end users are described from each pilot.

6.2 Pilot Specific Methods

In this chapter the currently used visualization technologies are described from all four pilots. In addition all pilots also describe the overall needs and identified requests from the end users. These are then used to identify the possible additional visualization technologies needed in a specific pilot during the first iteration. In addition these details are used together with the UX testing to define the further visualisations to be implemented at iteration two and three of the MIDAS Dashboard.

19 https://datavizcatalogue.com/methods/scatterplot.html

20 https://datavizcatalogue.com/methods/choropleth.html

6.2.1 Basque Country

Currently, stakeholders in the Basque Country use some simple and static graphs for analyzing the situation and to make decisions.

Public Health Manager / Director of Public Health and Addictions:

In order to present and report the public health situation in health commissions and to the public in general, the Public Health Manager / Director of Public Health and Addictions uses Microsoft Office Powerpoint presentations displaying public health status by means of line and bar charts.

Province-level Public Health Epidemiologist:

With the aim of analyzing the origin of health issues and to act on them, the Public Health Epidemiologist uses an Excel-based dashboard. These visualisations are mainly vertical histograms for incidence rate among gender per each age-range and tables showing the number of cases by zones and age ranges. In the table, the values that have unusual evolution are coloured in red. All this data is rate-adjusted by age. This dashboard is used in other health issues, but it is not available for children obesity yet.

Patient-group assigned pediatrician:

Pediatrician uses individual growth charts for tracking the physical evolution of their patients. Growth charts consist of a series of percentile curves that illustrate the distribution of selected body measurements (height and weight) in children.

These visualizations are quite limited. No difference has been found for analytics and visualization for people with different social status nor visualizations that allow the comparison of BMI with other variables.

The proposed visualizations are the following (per user story):

<u>Descriptive visualisations user story</u>

- Map-based visualisation with different aggregation level (Provinces, Regions vs Trust / OSI, municipalities, assigned patient-quota area) visualisation e.g. Choropleth
- Represent differences among regions.
- Represent positive / negative evolution in time.
- Rate-adjusted visualization
- Analytic applied to variable before being plotted.
- Line-chart with multiple lines, mark average and see which below, which above for different aggregation levels.
- Percentile graph for different aggregation levels.
- Scatterplot linked to map visualisation.

Variable correlation / relationship identification user story

- Correlation matrix
- Cluster visualisation
- Scatter-plot with PCA
- Radar-chart

Forecast / monitoring user story

- Short-medium-long forecast of variables representing child overweight and obesity
    - Use of the same visualisations as from descriptive and effectiveness assessment, but overlaying, following the line in a different colour for forecasts.
- Exploration of data model-based alternate scenarios, based on predictor variables options testing
    - Editor of sets of predictor variable configurations, that get plotted as short-medium-long forecast of variables' targeted visualisations
- Monitoring of implemented policies
    - Alerts within dashboard if current (or updated short-term prediction) child overweight and obesity deviations occur from the implemented policy forecast.
    - Visualisation through overlays and / or in lines with different colours of implemented policy forecasted, what happened until now and the new forecast, based on updated data.

6.2.2 Finland

Currently, visualization techniques are not widely used in stated stakeholder cases. While in some instances plot diagrams, pie charts, and population profile visualizations are being used, they are based on a limited amount of static data.

At the moment, the visualizations that are being applied are made with regularly available software packages like SPSS or MS Office suite. In some day-to-day cases, handmade scribbles and figures are often used as visualizations to convey data/message/knowledge from data collectors to upper layer decision makers. However, on the national level, a good set of visualization examples are provided by THL in SotkaNet[21] and other statistical services like TEAviisari, WelfareCompass[22] and Terveytemme[23].

21 https://www.sotkanet.fi/

22 http://www.hyvinvointikompassi.fi/

23 http://www.terveytemme.fi/

Based on the identified user stories and scenarios the following advanced visualization techniques are identified to be useful in the Finnish Pilot; either in national / regional scenario or in both scenarios, depending on the data on which the technique is applied. However, this list of visualization techniques is not exhaustive, rather it will evolve over time as the Finnish Pilot gets access to the data and variables requested from multiple sources:

- **Network diagram**[24]: Cause and consequence analysis is identified as a key element, as it should help policy makers if better visualized. Network diagram is a visualization technique that shows relationship between numerous variables. It can display the causal relationships, perhaps indicate correlation, most importantly it shows the the ripple effects of decisions being made.
- **Area graphs**[25]: Area graph is a useful visualization tool to display the development of quantitative values over an interval or time period. They are most commonly used to show trends. This type of visualization can be applicable for both the regional and the national scenario to reflect on longitudinal indicator data.
- **Stacked area graphs**[26]: Stacked area graph is a modified version of the area graph where longitudinal data on multiple variables are presented in the same visualization. In this way, trends of multiple variables can be compared and possible correlations can be pointed out.
- **Pie chart**[27,28]: Pie chart and donut charts are useful tools when making decision on resource allocation. Therefore, this visualization technique relates more to the regional scenario in the Finnish pilot.
- **Choropleth Map:** Choropleth Maps display divided geographical areas or regions that are coloured, shaded or patterned in relation to a data variable. This provides a way to visualise values over a geographical area, which can show variation or patterns across the displayed location. In the Finnish pilot, this technique can be used in both regional and national level. On the regional level, it could be used to portray area profiles to make it visually clear for regional decision makers for which locations in the region requires what type of attention. Similarly, on the national level, an overall national comparison can provide better understanding on specific issues.
- **Timeline[29]:** A Timeline is a graphical way of displaying a list of events in chronological order. The main function of Timelines is to communicate time-related information, either for analysis or to visually present a story or view of history. Especially in the Finnish national case, this technique can be used to

24 https://datavizcatalogue.com/methods/network_diagram.html

25 https://datavizcatalogue.com/methods/area_graph.html

26 https://datavizcatalogue.com/methods/stacked_area_graph.html

27 https://datavizcatalogue.com/methods/pie_chart.html

28 https://datavizcatalogue.com/methods/donut_chart

29 https://datavizcatalogue.com/methods/timeline.html

reflect how different variables have changed over time and their effect in the present. Timelines can be used in parallel with Network diagrams to dig deeper into the causal relationships, where timelines portray in detail about separate variables. From another perspective, timelines are a very useful tool for the Preventive mental health to show how the situation has evolved over the time, which helps policy makers with decision making. Besides network diagrams, other graphs can be combined with a Timeline to show how quantitative data changes over time.

The above elements together with the common elements decided for the first iteration of the platform can also support the system dynamic simulations. Typically they require visualization of dynamic hypothesis diagrams, causal maps, time series curve sets, sensitivity graphs, stacked columns and bars. The user interface commonly also contains graphical tables for inserting what-if assumptions of different scenarios. The actual needs for system dynamic modelling forecasting tools visualizations will be defined based on the first iteration experiences during the second and third iteration.

6.2.3 Northern Ireland

The stakeholders from Northern Ireland have identified the following specific kinds of analytics or visualisations for the higher level requirements for the first iteration of the MIDAS platform. These are planned to focus better on the second and third iteration.

- Longitudinal data analysis to look at changes over time and to track the cohort of children as they move into and around the care system. This kind of analytics is currently impossible with current methods. This would assist in gaining a better understanding of what services are working most effectively.
- Use of mapping with the ability to focus on small areas according to Northern Ireland Statistics and Research Agency (NISRA) geography[30]
- Dynamic cross-filtering of reports by using multiple visualizations and selection methods.
- Ability to load additional datasets as they become available (e.g. educational attainment).

6.2.4 Republic of Ireland

The most important focus from the Republic of Ireland point of view is to bring the healthcare data from numerous silo sources to a single access point. It can then support analytics and visualisations which enable data driven policy development and resource allocation based on population and trend analyses from evidence based data.

30 http://www.ninis2.nisra.gov.uk/public/documents/NISRA%20Geography%20Fact%20Sheet.pdf

Data visualisation takes many shapes in the Republic of Ireland. From a national perspective, Health Atlas[31] is a resource used which has open source data geomapped by hospital regions and community regions. Health Atlas Ireland is guided by Health Intelligence/Knowledge Management HSE in collaboration with numerous bodies.

The Irish pilot has a broad target audience and accordingly all common formats of visualisation should be provided to satisfy user requirements. The following data visualisation are identified as a priority:

- Data visualisation directories geomapped by hospital and community regions.
- **Choropleth Map** of the Republic of Ireland depicting:
  - Population distribution across Ireland
  - Deprivation patterns geographically urban and rural
- **Dot Density Map** of Hospital showing distribution of
  - Acute Care services
  - Primary Care
  - GP and Health Centres
- **Correlation charts** with **scatter plot** to show incidences of diabetes by employment rates and deprivation
- **Population Pyramid** to show distribution of diabetes Type 1 on Lt and Type 2 on Rt by age, region, comorbidity
- **Bubble graph** to show variables as above (age, region, comorbidity) with diabetes in a different format
- **Cumulative Curve** to show incidences of diabetes over a time period, e.g. a 10 year time span.
- Data Modelling and predictive analysis with forecasting tools to predict diabetes incidences, short term and in to the future
- **Percentile Charts** and run charts and **Ordered Bar Chart** to show budgetary allocation of resources (finance allocation and staffing location (Clinical Nurse Specialists)) related to diabetes compared to incidences of diabetes diagnosed for that region.
- Ranked lists of comorbidities/prescribing patterns per geographical regions
- **Stacked Columns** to show adverse events aligned to medication management of diabetes
- Chart depicting symptoms of diabetes relating to medication used to determine efficiencies of new medications methods
- Process Control chart to demonstrate that hospitals across Ireland are meeting diabetes care targets (Care of diabetic foot, retinal screening etc.)

31 https://www.healthatlasireland.ie/

- **Line Chart** depicting the number and frequency of hospital admissions per type of diabetes or complication of diabetes (amputation, vision issue, hypo or hyper situations).

6.3 Methods for External Analytics

Analytics and user interfaces in the MIDAS platform can be shared in two categories, internal analytics, which use the data dependent on the current pilot, and external analytics from global resources. This section describes the visualization methods connected to the external analytics for Social Media, Online News and MEDLINE resources. The key difference from a user interface perspective with the external analytics in comparison to internal analytics is that each of them are controlled and mainly used on its own dedicated dashboard and only the main results from them are brought into the main MIDAS Dashboard.

The social media analytics resources described in 4.5.1 provide tools for the end user to gain an overview for public opinion under investigated social media campaign and also an alternative option to apply traditional webforms based interviews or data collection studies. Together with dedicated Social media campaign management dashboard and MIDAS Dashboard they can provide new approaches to measure the impact and public opinion together with local data driven analyses.

The tools developed on top of data sources described in 4.5.2 and 4.5.3 are exploratory interactive modules that permit the user to explore Online News and MEDLINE. These tools are constructed based on previous research efforts and feedback from policy-makers at non-technical meetings and throughout project communication. The main goal is to adapt existing generic tools to public health usage scenarios and, additionally, to develop new visualisation modules where gaps have been identified. In addition to visualisation of text-related document analysis, possibilities of StreamStory technology are studied. The StreamStory tools may be integrated in the MIDAS dashboard during the iterations 2 or 3 based on the knowledge we learn from iteration 1.

6.3.1 Social Media Analytics

The results of the live user story used will determine exactly what kinds of visualizations will be used. The proof of concepts versions we are displaying in MIDAS prototype iteration 1 contain the following graphics:

- Custom progress bar / slider
    - Used to indicate the overall sentiment towards the policy (e.g. 0 = bad, 0.5 = ok, 1 = good).
- Pie charts

    - Used to show a break down of the percentage of negative, positive or neutral sentiment.
    - Used to show a break down of the percentage of Yes, No and maybe answers provided by the public.
- Bar charts
    - Used to show the percentage of various emotions contained in responses. (e.g. showing the amount of "anger" compared to the amount of "joy").
    - Used to display the results of multiple choice questions, where each value will have a positive and a negative bar.
- Map based visualisations
    - Current proof of concept extracts locations from Twitter profiles and displays where users are commenting from.
- Ranked lists
    - Lists ordered (decreasing) on the occurence of certain questions or suggestions being made by the public, to show policy makers the most frequently asked questions or suggestions.
    - Also used to display the most negatively talked about topics or persons.

A summary of the results will be displayed as a widget on the MIDAS Dashboard homepage. The full results and the campaign controls will be accessed through a link to a separate dashboard. The inputs needed for the social dashboard will be quite different to that for the rest of the system, for this reason they have been separated to avoid making the end tool too complex.

6.3.2 Online News Analytics

In the context of Online news monitoring, there are several views that can be used, depending of the analytical context. Taking into account the volume of news, which amounts to 150.000 news articles per day, a top down approach is used, which means that users are provided either with high-level summaries of prioritised list of topics. Afterwards, a drill down do a single news article should be allowed, in order to support informed decision making.

The following are the main groups of existing visualisation techniques, that will be extended and adapted to needs of public health policy making:

- **Real-time News Monitoring**. Real time status of world news, with highlighted the latest health related topics. The module also provides a search box where the user can look for a specific topic and have a visual representation of the worldwide news on that topic over more than 60 languages. The user can click on a particular news and this will open that particular news instance with title, summary, date and location, as well as number of shares in social media and list of related news.

- **Timeline News Navigator**. Data representation based on: (i) chart representing the number of articles per date for certain searched keywords, and (ii) a world map with number of articles per location during the selected time window. These permit the user to explore through time a certain Public Health related event based in the worldwide news published about it.
- **Storyline News Module**. Curve with publishing times of articles describing the chosen event where the height indicates the number of articles published in last x hours. In this module it is possible to explore the evolution of a single event based on the series of related news articles around it.

6.3.3 MEDLINE Analytics

Modules in this section build upon the MEDLINE dataset stored on a local ElasticSearch insurance and analysed by tools provided in WP4. The visualisation modules to be provided are as follows:

- **SearchPoint article prioritisation** - the user writes a few keywords in the search box that output: (i) a numbered list of 10 MEDLINE articles with a short description extracted from the first part of the abstract; (ii) a word-cloud representing the k-means clusters of topics in the articles that include the searched keywords; (iii) a pointer that can be moved through the word-cloud and that will change the priority of the listed articles.
  This visual iterative reasoning loop support helps surfacing information that one is looking for by redoing the prioritization in the search index output. It lets the user interact with the index of results by moving the SearchPoint pointer over the clustered topics represented as word-clouds.
- **Medline Analytics Explorer** - A dashboard that permits the user to explore data visualisation modules that feed on MEDLINE. Several pie charts, bar charts, etc, that can be moused over and provide information on several aspects of the MEDLINE data. In particular (i) the bar chart of counts of articles related with a certain topic; or (ii) a pie chart with the different concepts included in the articles that were provided by the user search.
- **Medline Atlas** - Based on the choice of a time window and a certain topic, such as childhood obesity, the user is able to view the clustered MEDLINE documents, mouse over the plot or click to view plotted points. It is based in the collaborative research with the MIDAS partner, Arizona State University. It aims to identify areas of dense scientific research corresponding to searchable topics and time windows. This technology will not be implemented in the first MIDAS prototype, but the needs are studied within implementation of other MEDLINE analytics.

7 System specification

MIDAS visual analytics interface for end users includes two main components and connectivity to external resources shared by all MIDAS Pilots. The main components are the flexible dashboard generator UI and middleware component which enables the UI to connect all necessary analytics interfaces. Several state of the art technologies are used to enable all functionalities on the MIDAS visual analytics tools. Figure 7.1 describes the components used to create the MIDAS visual analytics tools and this chapter describes the main approaches used to implement these on the first iteration of the MIDAS prototype.

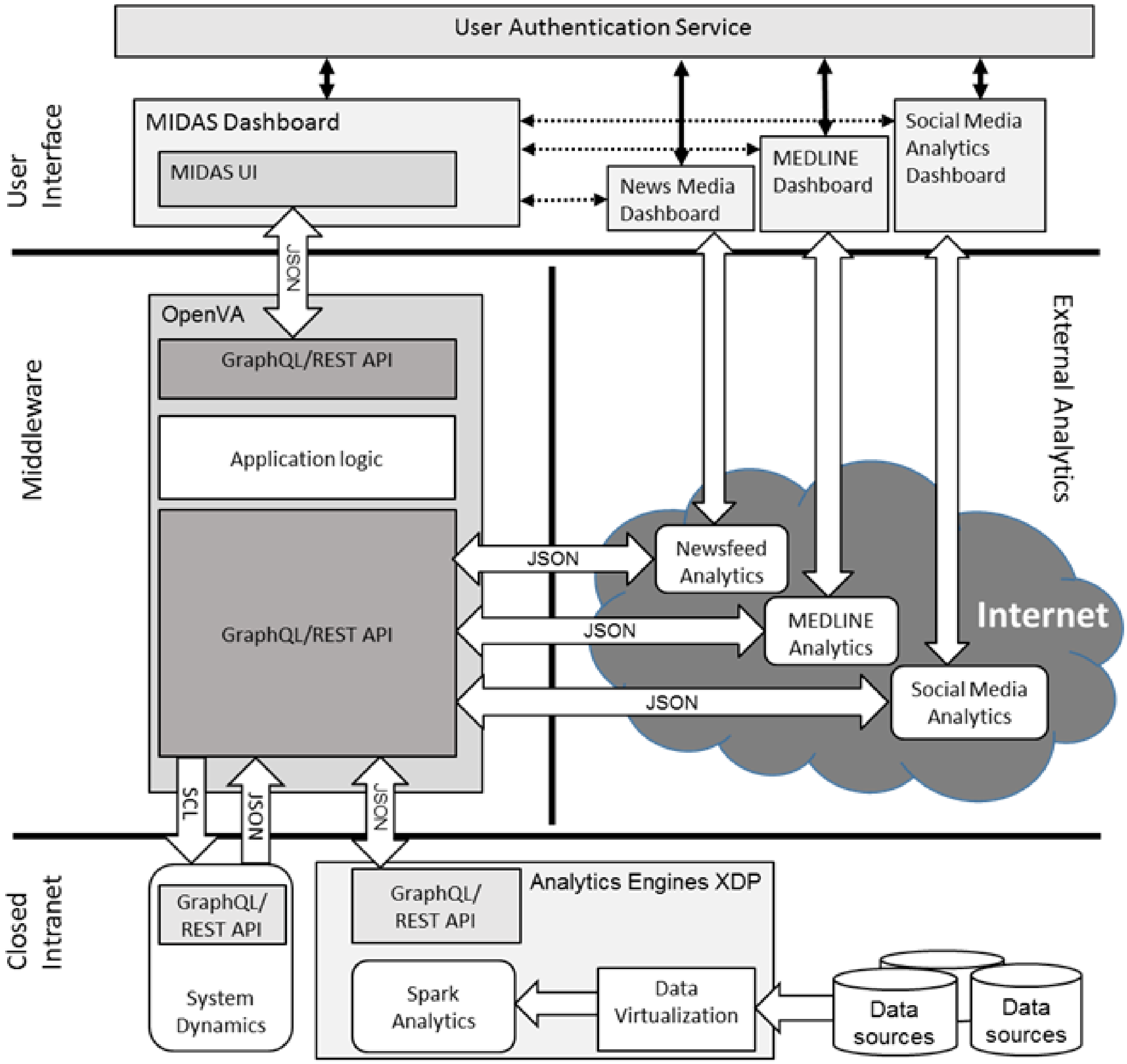


Figure 7.1: MIDAS Dashboard concept. The MIDAS Dashboard includes two main components: The MIDAS UI and OpenVA based Middleware component. The UI is communicating with the end user and the OpenVA is responsible to communication between UI and all analytics services used in MIDAS concept.

7.1 User Authentication

MIDAS system has two user authentication schemes: The researcher authentication on backend services and end user authentication on frontend services. This document considers here only the frontend services authentication for the end users. The final implementation of MIDAS front ends will share a single sign-on service, which is utilized over all end user dashboards:

- MIDAS Dashboard,
- Newsfeed analytics dashboard,
- MEDLINE analytics dashboard, and
- Social media analytics dashboard.

The basic principle is that the user does not need to authenticate more than once. After a single authentication session the user can access each of dashboards listed above. The dashboard users should access only results of the data analytics in the MIDAS dashboard and they should not have access to the actual data or raw analytics. For this reason the researcher authentication for backend services is behind separated authentication services.

In practice, when user wants to access any of the four dashboards the access rights are checked from the authentication service. If the user is not yet authenticated they are redirected to the authentication service and after successful authentication, the targeted service can be accessed. After this, the user can access any of the MIDAS dashboards without new authentication as long as the authentication session is alive. On the first iteration the authentication service is implemented only with the main MIDAS Dashboard and the best practices to utilize it for the external dashboards are studied with knowledge learned from the first iteration.

One challenge with the single authentication service is that each MIDAS pilot deployment is located in systems with somewhat restricted Internet access. Each of them will handle the authentication on their own locally, but they all use the same external dashboard services for Online News, MEDLINE and Social Media analytics as described in Figure 7.1. The knowledge and experiences from the different deployments of the pilots will provide the information required to select and design a secure common authentication service which can utilize and serve all relevant parties in MIDAS system.

From the initial review of the current authentication technologies the OAuth 2.0 Framework has been selected for the initial approach to create the authentication server. Clients from each dashboard environments will then use the server. During the first iteration Online News, MEDLINE and Social Media analytics dashboards are using separated services, but they collaborate tightly to develop the common authentication service. The OAuth 2.0 Framework also supports the use of external authentication services which usage is studied separately on each pilot deployment. These kind of services can be, for example, the authentication systems of partnering universities or possible national authentication services like HAKA authentication in Finland. If the security agreements allow, Google OpenID, Facebook Login, MS Azure AD Login or other similar services can also be considered. These also enable multi-factor authentication schemes with additional media like SMS.

7.2 User interfaces

The end user interfaces of the MIDAS system includes four components:

- MIDAS Dashboard
- Social Media analytics Dashboard
- NewsFeed analytics management Dashboard
- MEDLINE analytics management Dashboard

The MIDAS dashboard is the main user interface of the system and the three other dashboards are serving the management interface for the complex analytics needed with the Social Media, Online News and MEDLINE resources. The results of these management dashboards can be linked to the MIDAS Dashboard with a widget system. The common user identification interface serves all four dashboards.

7.2.1 Reasoning Loop Support

The user interfaces of the MIDAS Dashboard are aimed to support reasoning loop workflow. The concept of reasoning loop support is visualized on Figure 7.2 below. The main idea on reasoning loop workflow is that user can easily modify and save the applied analysis and visualisations based on the findings they identifies from those. It is preferred that the user can also return to earlier results or states if needed. This creates certain basic requirements for the user interface.

Supporting reasoning loop approach requires that user interfaces enable users to re-select the data, analytics types and visualisations. In practice, these can be achieved by enabling certain functionality on the dashboard. Some of these fundamental functionalities are:

- Option to add, modify and remove visualizations
    - Option to re-select the used data
    - Option to change the parameters of the visualisation and analytics
    - Option to change the selected analytics and visualisation for the data
- Option to save the current dashboard by new name
- Option to modify the structure of the dashboard

Not all of these are needed at once and many of them affect the others. For example, by changing data, the user must re-select the analytics and by changing the analytics, the user must also re-select the visualisations as new selections might not be compatible with previous ones. One key concept with reasoning loops are that new analytics can be easily generated from the previous figures.

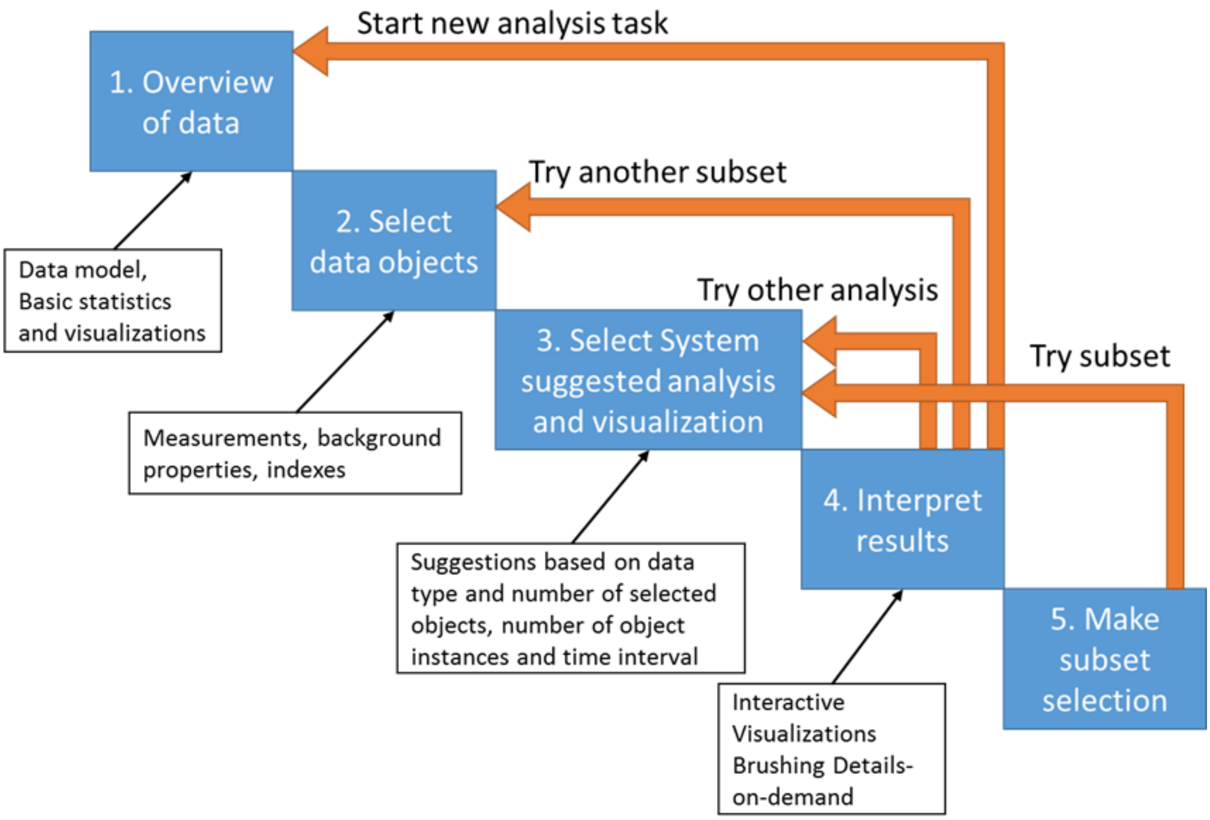


Figure 7.2: Reasoning loop logic. The MIDAS Dashboard will support end users to use reasoning loop logic when studying data and creating dashboards. The key concept with reasoning loops is to possibility to modify the earlier selections according to the detections from them.

7.2.2 MIDAS Dashboard

The MIDAS dashboard application provides a user-friendly interface to create dynamic dashboards powered by strong data analytics engines supporting the reasoning loop approach. It allows different stakeholders to access various features and functionalities without the danger of exposing sensitive raw data for them. The idea of dynamic and flexible dashboards is achieved by using the concept of a widget. Each widget in the dashboard depicts a graphical or textual representation of data received from the OpenVA server. The widget serves as a separate entity within the dashboard and end-user can add, remove, update, resize or reposition an individual widget independently with respect to the other widgets on the same dashboard. The widgets are responsive and resize dynamically based on different screen sizes.

Grant Agreement No: 727721

Adding a new widget requires selection of datasets, various parameters and a type of visualization. To facilitate this process, the dashboard application provides a wizard, which allows the end-user to select inputs for the new widget. The widgets can include one or more figures and certain figure types are able to trigger the wizard to generate new widgets.

The end-users can create multiple dashboards according to their needs. All dashboards created by the end-user are saved into a database and are accessible during the future sessions. As dashboards are fully customizable, the end-user can add or remove widgets at any time, and can update the name of the saved dashboard. The owner of the dashboard can also share it with other users of the dashboard application. Figure 7.3 illustrates the initial dashboard interface including three widgets with different data sources.

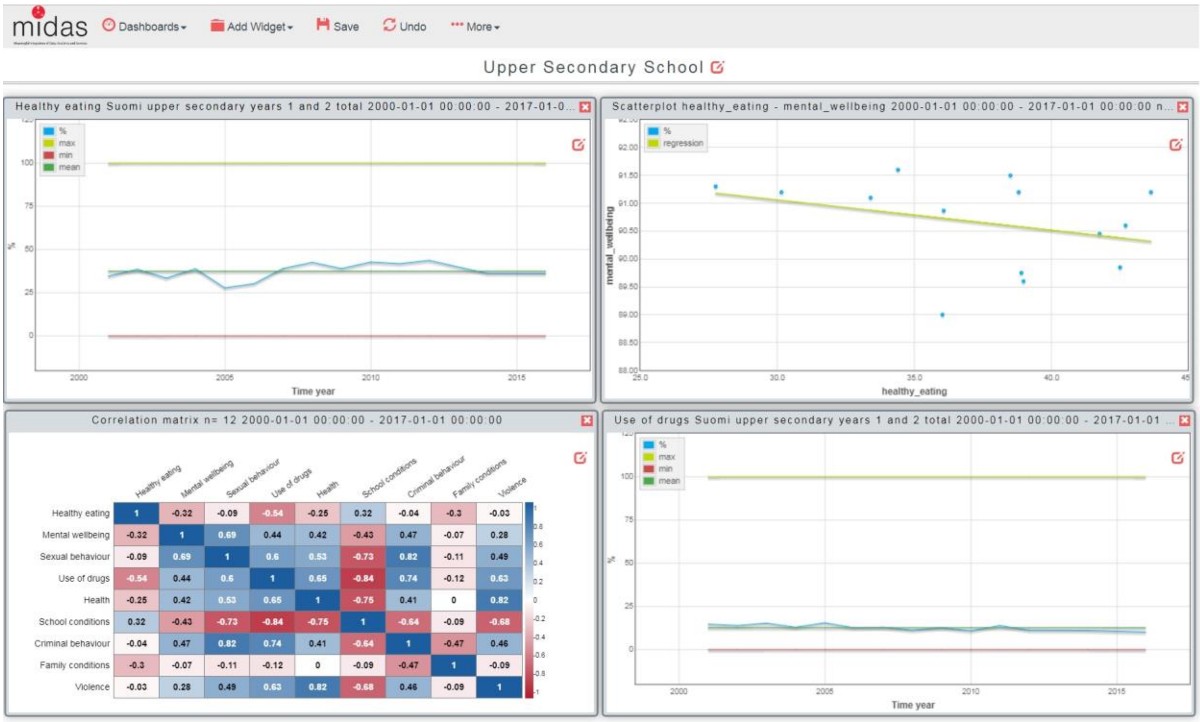


Figure 7.3: MIDAS Dashboard demonstrator. Screen capture from current mockup prototype version of MIDAS Dashboard (generator). In this example dashboard, the user has added four widgets with different variables. Top-left and bottom-right widgets are line graphs as a visualization method. The top-right widget has a scatter plot where a regression line has been fitted and visualized together with the data. The bottom-left widget includes heatmap presentation of correlation matrix over various variables.

The main functions of the MIDAS dashboard are:

- Add/modify/replace graphs/analytics,
- Share generated dashboard to other users,
- Save and make copies of dashboard,

- Undo the most recent action,
- Include interactive graphics on the widgets
    - Values can be shown by mouse click,
    - Connected figures where selection in one updates others
    - Option to generate new widgets from the selection of (certain type) of graphics
- Export graphs from the Dashboard.
- Combine statistical analysis with social media, news and MEDLINE studies.
- Use forecasting tools to support preventive policies and policy effectiveness assessment.
- Maintain the security of the data
    - On the generation phase end users see only datasource and variable names, no actual data.
    - With widgets only to the visual representation of the results computed from the selected variables are shown.
        - Showing too small results will be restricted.

7.2.2.1 Technology

The core web-technologies used to develop the prototype for the first iteration of MIDAS Dashboard user interface (UI) framework are the HTML5 (Hypertext Markup Language), CSS (Cascading Style Sheets) and JavaScript. In order to keep the UI lean, robust and scalable following frameworks and libraries are taken into use:

- **jQuery**: (version 3.2.1) is a JavaScript library that allows easy and fast manipulation of the DOM (Document Object Model) elements in the HTML document. It also provides functionality to interact with external REST APIs by using AJAX (Asynchronous JavaScript And XML) and handles various other important tasks required to develop a web-application.
- **Bootstrap**: (version 3.3.7) is a CSS framework for designing and development of modern and responsive front-end of the web applications.
- **Gridstack**.js: (version 0.3.0) is a jQuery based plugin that allows developing responsive widget based layouts. It is also compatible with the bootstrap framework.
- **Moment**.js (version 2.18.1) is a JavaScript based library that handles all the complex operations relevant to date and time.
- **flotr2**: (version 1.2.0) is a JavaScript based library for drawing HTML5 based graphs. It provides capabilities to draw most of the simple and complex visualizations.
- **Leaflet:** (version 1.2.0) is a JavaScript library for creating interactive choropleth maps with the help of GeoJSON data format.

7.2.2.2 Widgets

The MIDAS Dashboard is designed to utilize widgets which include selections for:

- Data
- Analytics
- Visualization
- Parameters for the Analytics and Visualization

The widget is added to a dashboard by using a widget wizard. The wizard includes multiple stages to define the needed information. The Figure 7.4 presents the basic flow of the wizard.

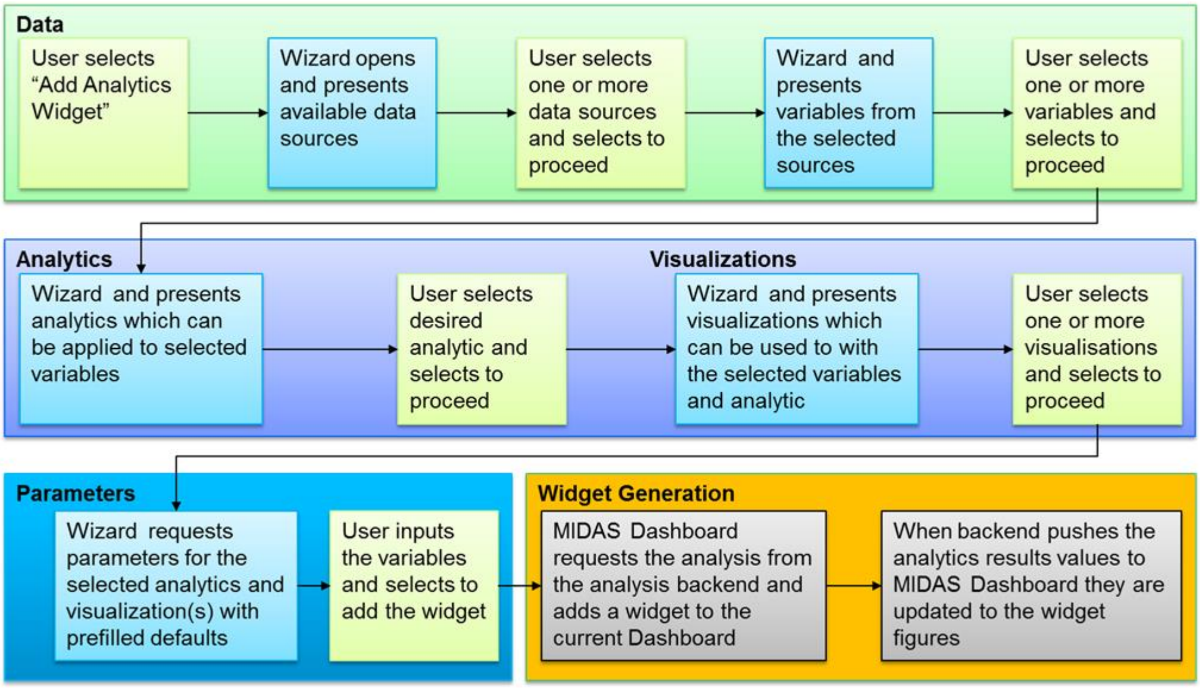


Figure 7.4: Widget generation wizard workflow. The initial definition for MIDAS UI Widget wizard. At first phase user selects data and variables and based on those wizard will show possible analytics for them. Based on analytics wizard shows possible visualisation methods to be selected. At last phase parameters for the analytics and visualizations are requested according to the earlier selections. After this the system draws the widget and updates it when the data is available.

The functionality of wizard can be described in three phases. First user selects data and variables, then analytics and visualization and at last defines parameters for the analytics and visualisation. After these are verified, the wizard triggers the needed processes to the middleware and generates an empty widget where the figure is generated when the analytics system returns the numerical results.

As the figure 7.4 shows, the selections are dependent from each others. The available analytics depends on the selected data and variables, and the possible visualizations depend on the used analytics. Data might need parameters like time scale, area or age group. Most of the analytics have their own parameters as well as have the visualisations. All of these are handled with metadata model used by

OpenVA middleware component of MIDAS Dashboard, which is updated according to the available data, analytics and visualisation methods on the current system.

In practice, the widgets are JavaScript functions which draw the results on the screen. To enable all MIDAS partners to develop these widgets a standard specification for the widget structure will be defined for the first prototype. The standard specification includes the input and output formats for the widget data and how the drawing environment should be used. The initial version of the custom widget requirement document is on Appendix 2.

In addition to common widgets there will be two special widget types. One for system dynamic modelling and one for external analytics. Both of these cases are not computed on MIDAS analytics platform but on separate servers. The communication between these and MIDAS Dashboard is handled by OpenVA framework and if possible using the GraphQL functionality. These will follow the guidelines of custom widget requirements but will include additional fields and functionality in comparison to the typical case.

7.2.3 Social Media Dashboard

7.2.3.1 Technology

For the frontend, IBM is using a similar stack to Analytics Engines:

- **Angular: (version 2.4.1)** is a framework developed by google for building single page web applications which include server side rendering for increased performance. Google skipped version 3 and released version 4, which we will look to use for the final version of the dashboard.
- **Bootstrap: (version 4 alpha 18)** is a CSS framework for designing and development of modern and responsive front-end of the web applications.
- **Webpack: (version 2)** is a tool used to compress, minify and aggregate javascript, css and assets reduce the size and load time for opening the application.
- **D3: (version 4)** is a javascript library used to draw custom charts, visualisations and animations.

The backend will consist of:

- **Node.js: (version 6.9.5)** is an open source javascript framework used to develop scalable server side / cloud based applications.
- **Express: (version 4)** is an open source Node.js module used for building REST based API's. IBM will investigate using GraphQL in later iterations.
- **Cloudant:** Cloudant is IBM's hosted version of the open source technology CouchDB, with additions to make it more suitable for big data solutions. Cloudant is a NoSQL based used to store JSON documents and run complex searches with Apache Lucene.

- **REDIS:** is an open source in memory database designed to store small amounts of data, but with an extremely quick response time. It is frequently used for session token or state management.

As an initial phase the authentication is provided by Superlogin (version 0.6.1), which is an open source OAuth2 middleware for Node.js and it is built by a contributor to the Angular project. Superlogin uses Cloudant (or CouchDB) to store the user details and REDIS to manage sessions. It also provides a full REST API and has a plugin architecture to enable third party authentication solutions such as Facebook and Google+. This will be replaced by the MIDAS login service when it is available.

7.2.4 News Media dashboard

7.2.4.1 Technology

- Frontend
  - D3.js
  - Angular
- Backend
  - Node.js
  - MongoDB
  - NewsFeed (http://newsfeed.ijs.si/) - A clean, continuous, real-time aggregated stream of semantically enriched news articles from RSS-enabled sites across the world. Uses PostgreSql for the database and provides REST API written in Python. Developed at Jozef Stefan Institute.
  - Wikifier (http://wikifier.org/) - a web service that takes a text document as input and annotates it with links to relevant Wikipedia concepts.
  - REST API written in Python. Developed at Jozef Stefan Institute.
- Authentication
  - Oauth2

7.2.5 MEDLINE analytics dashboard

MEDLINE articles are stored in an ElasticSearch instance hosted on QUINT servers. ElasticSearch provided REST API to modules listed below. Those modules have their own backend that uses the provided REST API.

7.2.5.1 SearchPoint

- Frontend
  - KineticJS for drawing and interacting with the widget
- Backend
  - Node.js
  - QMiner for analytics
- Authentication

- Oauth2

7.2.5.2 Medline Analytics Viewer

The definition of this module is still in an early exploratory phase. Kibana dashboard will be used for testing and prototyping. The technology used for the development of the final version will be selected in a later phase of the project.

- Frontend
  - Kibana for prototyping
- Backend
  - ElasticSearch
- Kibana for prototyping

7.2.5.3 MedlineAtlas

- Frontend
  - D3.js
  - CSS
  - jQuery - a cross-platform JavaScript library designed to simplify the client-side scripting of HTML.
- Backend
  - QMiner (http://qminer.ijs.si/) - a data analytics platform for processing large-scale real-time streams containing structured and unstructured data. Developed in collaboration with Jozef Stefan Institute. (http://qminer.ijs.si/) - a data analytics platform for processing large-scale real-time streams containing structured and unstructured data. A product of Jozef Stefan Institute.
- Authentication
  - Oauth2

7.2.6 Stream Story

7.2.6.1 Concept

StreamStory is an exploratory data mining tool based on a system for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behavior of the systems' dynamics. It provides an interactive diagram representing the cycles in the dataset, based on a sequence of states of different size (representing time spent in them) and relation between states. This includes choice of parameters, timescale or clustering algorithm to use by choice on dropdown menus. Moreover, this permits us visualisation modules, such as in Figure 7.5 representing the amount of rain in the UK as an example.

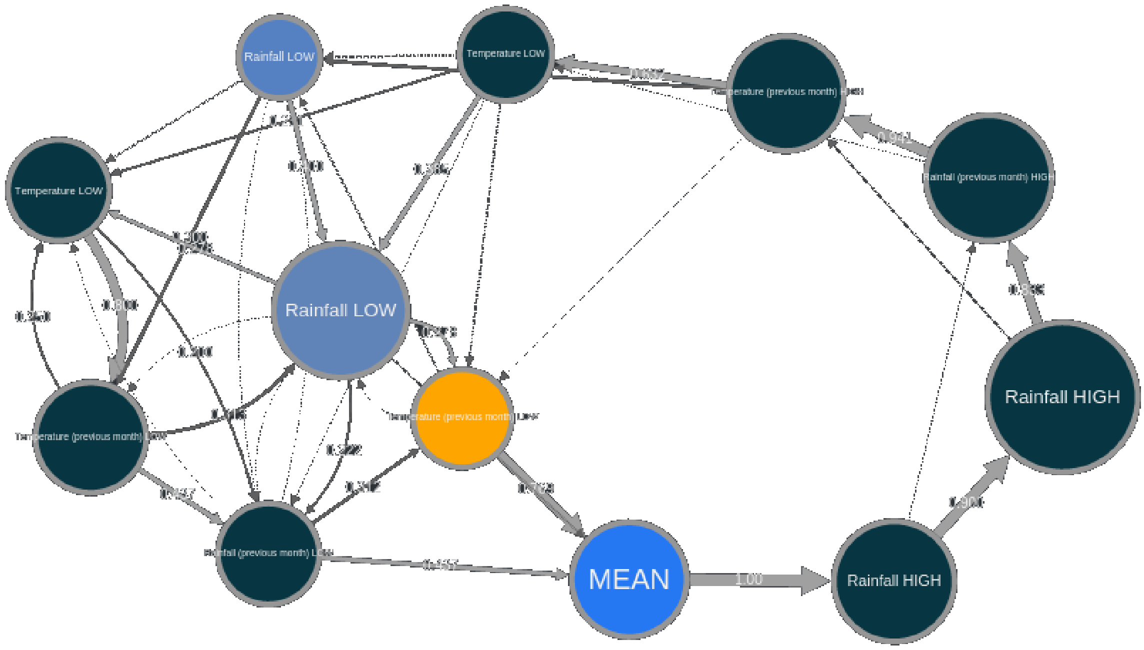


Figure 7.5: StreamStory visualization module. An illustrative example of how StreamStory represents periodicity in the data, identifying the time spent in each state and all state transitions.

This technology will not be implemented in the first MIDAS prototype, as its true potential is still being explored with the MIDAS datasets that are now available. This will be developed alongside the duration of the first prototype, and will be improved by what is learnt from the user experiences of the first implementation, which can then be used when it is implemented on the second round of the MIDAS system.

7.2.6.2 Technology:

- Frontend
  - Bootstrap for design
  - jQuery to interact with the DOM
  - Cytoscape for the central graph-based visualization
  - D3 for histograms, parallel coordinates and other plots
- Backend
  - Node.js
  - QMiner for analytics
- Authentication
  - By default StreamStory uses a custom authentication module. This module can however be replaced by most standard authentication systems.

7.3 Middleware technology

The middleware technology in MIDAS Dashboard has two main responsibilities:

- Handling the metadata to enable smart logic for UI (OpenVA logic)
- Apply efficient communication between UI and different analytics (GraphQL)

7.3.1 OpenVA

OpenVA[32] acts as a gateway between Midas UI and different other services like Simantics System Dynamics server[33] and MIDAS analytics layer. It enables REST/GraphQL interface for MIDAS UI to communicate with and REST/GraphQL for communication towards analytics systems. This enables feasible uniform access for Midas UI to different kind of data sources and services but also implementing other kinds of Midas UI in the future like native UIs for some dedicated mobile environments.

The OpenVA is developed by VTT with Spring Framework[34] using the following projects:

- Spring Boot to allow quick development on production ready application.
- Spring Data to internal data access of metadata.
- Spring Security for protecting the application with authentication and authorization.

Developing with Spring is done with Java which also allows native utilization of JavaScript with Rhino JavaScript Engine if needed. Handling metadata with OpenVA brings additional value for utilizing the other services. Eg models of System Dynamics contain several values but only some of them are in general modified. As a result OpenVA can handle storing the different views and access to the model independently of the model execution.

OpenVA is handling the metadata information with a specific metadata model, which is updated according the available data, analytics and visualisations on the environment. In the MIDAS pilots the OpenVA metadata model is Pilot specific. The initial metadata models are defined for the first iteration and during the development the options to update the metadata model automatically are studied. A reference of a typical OpenVA metadata model is found from Appendix 3: Conceptual OpenVA Data Model. The actual MIDAS Dashboard communication model will be included in D5.2 Visual Analytics Tool concept V2.

32 https://www.thinkmind.org/index.php?view=article&articleid=infocomp_2013_3_10_10030

33 http://sysdyn.simantics.org/

34 https://spring.io/projects

7.3.2 GraphQL

GraphQL is an advanced query language developed by Facebook. It enables efficient fetching of data with heterogeneous environments where traditionally REST[35] servers are used. Both REST and GraphQL[36] approaches can be used with URL requests and both of them can return JSON as a response. There are still fundamental differences between these two. The key difference is that with REST, the server determines the response, but in the GraphQL, the client makes the decision what kind of response it needs.

The GraphQL also enables fetching related resources with one query compared with the REST where several general requests are usually made. While REST combines the type of resource and how the resource is fetched with the REST endpoint, the GraphQL decouples these concepts. GraphQL describes them separately with schema, which contain types that describe the resources and queries, which describe how the resource can be fetched. Instead of endpoints, GraphQL provides resolvers, which matches the fields of types of schema and which can be called several time during the fetch.

Below is an example of GraphQL schema that can be utilized as a basis for resource handling in system with GraphQL:

```
type AttributeValue{
      type: String
      floatValues: [Float]
      txtValue: String
}
type Attribute{
      name: String
      attributeValues: [AttributeValue]
}
type Thing{
  name: String
  attributes: [Attribute]
  things: [Thing]
}

input AttributeValueInput{
      type: String
      floatValues: [Float]
      txtValue: String
}
input AttributeInput{
      name: String
```

35 https://en.wikipedia.org/wiki/Representational_state_transfer

36 http://graphql.org/

```
      attributeValueInputs: [AttributeValueInput]
}
input ThingInput{
  name: String
  attributeInputs: [AttributeInput]
}
type Query {
  metaData: [Thing]
  dataModel(input: [ThingInput]!): [Thing]
}
type Mutation{
  addThing(input: ThingInput!) : Boolean
}
```

8 Implementation Plan

8.1 First iteration

This document describes the technology and methods which are taken into use with MIDAS prototypes with three iterations of the implementation. This section defines the technology which will be implemented at the first iteration of MIDAS platform within MIDAS Dashboard. The research and development is not limited to the following technology but they are having the main focus on the iteration and additional technologies are possible if time and resources enable them.

8.1.1 MIDAS Dashboard

The key technology which will be implemented are

- Limited support for reasoning loop type of user experience
- The visualization types described in section 5.1
  - Line graph
  - Bar graph
  - Scatter plot
  - Heat map
  - Choropleth Map
- Limited usage of GraphQL on communication between analytics and OpenVA framework. The OpenVA uses internally GraphQL, but most of the analytics endpoints are implemented at REST API basis.
  - Study for transforming these to GraphQL is applied during the first iteration
  - Efficient communication is implemented between OpenVA and Analytics Engines XPD system.
- Limited implementation of Pilot dependent visualizations
  - All pilots will have localized geographical map based visualisations
  - Finnish pilot will have limited user interface and visualization for System Dynamic Modelling services
- The end user authentication will have initially limited implementation with OAuth 2.0 framework. Only main MIDAS UI will use the common authentication service. During the first iteration the external analytics are using their own authentication services which are implemented also with OAuth 2.0.
  - The best approaches to unify the authentication on second and third iterations are studied during the implementation and user experience testing of first iteration.
- The MIDAS Dashboard is deployed on each pilot side independently with needed localization on the system.
- Simple user feedback and experiences are collected during user experience evaluation with first iteration MIDAS platform implementation.

8.1.2 Social Media Dashboard

- **Authentication** - The dashboard will have its own user authentication, allowing registration and login to all users. Dashboard will whitelist email domains from consortium members and policy board to ensure no unauthorised users access the dashboard.
- **Ability to run multiple campaigns at once** - The dashboard will only have one public facing twitter account, but that account will be able to respond to the public about multiple campaigns at once.
- **Display basic results of questions** - In order to deliver quantitative results, similar to those seen in online survey tools, such as survey monkey, the dashboard will display pie charts and bar charts showing the basic results of each question. This means highlighting how many “Yes”, “No”, “Maybe” or “I don’t know” responses were given as responses by the public.
- **Display a small number of deeper insights** - For the initial release, the dashboard will display a small number of deeper insights, going a level beyond the basic results. These types of results will be similar to highlighting the most common suggestions that the public have made for the given policy, gathering and displaying responses to multiple choice questions that were not included in the options given to the public. These types of results will be more valuable to the policy makers as they will help to identify issues and concerns that the public have with the policy.
- **Chatbot logic and questions controlled by a configuration file** - In order to facilitate multiple campaigns running at once, the system will need to have configuration files for each campaign that will allow the system to handle the necessary business logic. For the initial release these will be created by hand, until a tool can be developed to provide the necessary options to the policy makers.
- **Chatbot will be able to answer limited questions from the public** - As highlighted from our initial POC’s, when the chatbot is asking questions, typically the public will need to ask a question back to understand what is being asked. For the initial release, the chatbot will have a limited ability to answer some of these questions.

8.1.3 News Media Dashboard

- Global monitoring of cross-lingual worldwide news based on public-health related queries. This is represented over a world map with pointers to the location of origin of the news articles mentioned (location extracted from the text or publisher broad location, if the latter not available). The news articles include title, first paragraphs, publisher, publishing date/time and social media indicator (counting the number of Twitter shares).

- Timeline of events based on a bar chart representing the amount of news on a particular topic over time. This permits the user to explore published news articles around events that match the search criteria.
- Event categories represented by a pie chart showing the distribution of events into different categories. When the user clicks a particular segment, it displays events in the category and its percentage in the news assigned to the query.

8.1.4 MEDLINE Analytics Dashboard

- Prioritization of MEDLINE abstracts of articles displayed as result of a query based on keywords. The user explores that prioritization with the movement of a pointer over the clustered concepts extracted from the articles listed, represented as word clouds.
- Exploration of quantitative aspects of the MEDLINE dataset through interactive data visualization modules. The latter include bar charts, scatter plots, pie charts, etc, appropriate to the data they are representing.

8.2 Future iterations

Some methods and technologies are identified to require more knowledge and hand on experience from the implemented system for which reason they are now planned to be under research for the first iteration of MIDAS platform. They will be taken into action during the second and third implementation based on the learned knowledge from the first iteration.

The most important ones for these are:

- Streamstory visualisations and analytics
- Advanced functionalities of the dashboard to support the reasoning loop analytics
- Common user authentication implementation support over all end user interfaces

8.2.1 MIDAS Dashboard

For the main MIDAS UI there are already identified certain functionalities which will be under investigation and research towards second iteration:

- The widget generation wizard working order through data, analytics and visualization selections
    - Developments of the wizard are tightly connected to the user experience testing of the first iteration implementation
- The needed advanced visualisation methods and graphs in addition to advanced interaction within the graphs

- Methods for effectiveness assessments and feedback from the user experience testing are defined in detail based on the experiences learned with the first iteration user experience testing of MIDAS platform

With the OpenVA and GraphQL implementation the key topics which aim towards second iteration are:

- Pure GraphQL communication between Analytics Engines XPD based analytics
- GraphQL based communication with external analytics resources
- Optimized communication with analytics resources
    - Buffering of the calls
    - Bundling the calls
- Methods and implementation for automated OpenVA metadata model updates from all analytics resources.
- Robust and efficient usage of Simantics System Dynamic server resources

8.2.2 Social Media Dashboard

Known plans for the second and third implementations details

- **Ability to create a campaign through a page on the dashboard** - The dashboard will provide a screen where policy makers can create their own campaigns. This screen will generate a configuration file and deploy it into the application.
- **Ability to edit a running campaign** - While a campaign is running, the system may surface information that the policy maker will wish to address. For example, the chatbot might highlight a large number of questions being asked, that it can't provide an answer to. At this point the policy maker should be able to edit the configuration file to provide answers to these questions, so that chatbot can answer the public.
- **Other social media sources** - For the system to be effective, it needs to reach as many users as possible. In iteration 2 we will look to add Facebook support first, and then time depending move onto Reddit, Linkedin, Google+ and Instagram.

- **Multi-lingual support** - We will aim to provide a solution for the chatbot to be able to talk in multiple languages. Interest has been expressed by the university of OULU (finland) and Bioeff (Basque country) to use the chatbot. To accomplish this it will need to speak Finnish and Spanish.
- **Additional deeper insights** - We aim to provide deeper insights to the dashboard for future iterations, allowing policy makers to get the most out of the system.

- **Improved ability to answer questions** - We aim to improve the ability of the chatbot to have a conversation with the public, this will require improving the ability for it to answer questions from the public, as well as asking its own.

8.2.3 News Media Dashboard

Known plans for the second and third implementations details

- Concept graph representing the relationships present in the news corresponding to the query made by the user. The nodes in the graph are the most relevant concepts in the events while, the edges are displayed between the concepts that frequently co-occur in the same events.
- Clustering of events represented by a dendrogram displaying how events can be split into subgroups based on their similarity. Clicking on the leafs of the tree will display sub-clusters (to optimize the screen space).

8.2.4 MEDLINE Analytics Dashboard

Known plans for the second and third implementations details

- Mapping of MEDLINE articles in a bidimensional space, with mouseover information pop-ups and two-click redirection to article URL.
- Automatic annotation of free text, loaded in a webapp text box. This annotation is based in the MeSH terms (descriptors) recurring to a classifier learning over the MEDLINE dataset.

9 Discussion

Work towards this Deliverable had some unexpected challenges on user requirements collection work, which all were finally resolved. To conquer those the Deliverable was postponed to get all necessary details gathered and included to it. The main challenge was identified in the large heterogeneity of policymakers within and between pilots. Some policymakers were fairly well aware of the data visualization technologies and possibilities of how the multi-domain data can support policy making, but on the other hand some had only a limited know-how on these topics. This can be noticed from IT and Internet skills as well as data analytics skills of personas templates. This caused a significant contrast to user stories where in extreme cases policymakers were relaying handwritten textual reports and only simple spreadsheet tables and graphics if not even that. The aim of living document of Pilot descriptions was to collect a detailed overview for the technical team that which kind of cases the MIDAS Platform must be able to serve. Besides the personas exercise, a separated collection of guiding questions were developed to collect the information in mode coherent manned from every pilot area. The descriptions of pilots in this document are based on the content of the living document state at the end of October 2017. The living document is updated throughout the MIDAS project and it will support all deliverables, which require the details of the study topics or end user requirements. On MIDAS the next significant update for Living document will be at the end of August 2018 to support deliverable D5.2.

The WP5 related KPIs for user requirements at the first iteration were reached well at cross-pilot level. All pilots were able to define the research question and multiple scenarios within the research topic. The most critical visualisation methods were possible to identify from all pilots as well as the basic needs of end-users for the UI. We were also able to identify the technology to use for the implementation of the first iteration of MIDAS Dashboard. The more detailed needs and use cases will be addressed at D5.2 with the help of user experience testing at D5.6. All pilots were also able to define more than initially required four visualisation types. These will be addressed at later iterations of MIDAS Dashboard concept.

The main reason to move certain analytics and visualizations connected to them to be implemented only after the first iteration was the heterogeneous knowledge about analytics and visualizations of policy makers. This decision enables us to gather user experiences from the policy makers and also educate them during the first iteration of user experience testing connected to the task T5.3 and deliverable D5.6. Good examples from this are the StreamStory technology and advanced methods to support reasoning loop approach on analytics. Also some of the topics in the Grant Agreement were not applicable due to the high variance of end users and thus those were considered to be answered at the follow up deliverable D5.2 and user testing deliverable D5.6. These topics were the end user feedback and effectiveness assessment of user experience testing.

The heterogeneous knowledge of policy makers was not the only challenge within the differences between the pilots. Another identified challenge was noticed on deployment locations which all required significant research to find and plan making to cover all legal and privacy issues related to the system deployment. System deployment is mainly handled in WP3, but it has also effects for WP5 as the frontend security and authentication must be designed so that it can connect also to external resources and not only resources within the same network. In D5.1 this is handled by use of GraphQL and OpenVA as a connection point between all analysis resources. By this way the sensitive analytics can be located on isolated network where only communication between OpenVA and Analytics Engines XPD is allowed leaving the frontend and end user authentication to environment where more freedoms in connectivity are possible. This is not required on all pilots but it can be used in a similar manner within each of them. In the future iterations the common authentication scheme of end-users is studied as it can provide significant improvement on user experience.

10 Conclusion

This deliverable describes two main concepts within MIDAS project. It describes the pilots involved for the real life testing of MIDAS platform and it defines the main concepts for UI and data visualization for the first implementation of MIDAS Dashboard. There has been certain challenges to collect the necessary information for the topics, but all of them were successfully handled. As an unexpected event these challenges required the deadline of the deliverable to get postponed slightly to maintain the good level of the knowledge. The delay should not have a significant effect on the actual system development as the implementation work of the system framework has been done parallel to the user requirement definition work. This deliverable forms a solid base for the MIDAS Dashboard implementation at D5.3 and a good starting point for the follow up deliverable D5.2.

11 Glossary

Pilot is a solution (e.g. service or program) that has been done as an experiment or to test before being introduced more widely. Pilots are used for demonstration or for testing audience reactions before the full implementation of a product or a service will be out in the markets for public access or for full use.

Policymaker is a person responsible for or involved in formulating policies, especially in politics. In health politics, policy-makers are people who are involved in making policies and policy decisions of healthcare organizations. They are also responsible for making new rules or laws, and they are also making hard choices within regional or national health care priorities.

Scenario is a brief description of the local policy context in which the software is being used, and out of which the user story emerges. The scenario might include relevant political, environmental or social factors, specific policy or reporting requirements, or other challenges relevant to the use of the platform.

Stakeholder is a person (or group) owning a significant percentage of a company's shares, and/or a person (or group) affected by or having an interest in its operations, such as the employees, customers, local community, etc., or relating to policies intended to allow people to participate in and benefit from decisions made by enterprises/organisations in which they have a stake a stakeholder economy.

User Story is a short, informal description of a feature, generally told from the perspective of the person who desires the new capability, usually a user of the system. According to UK Government Digital Service, a user story should always include:

- the person using the service (the actor)
- what the user needs the service for (the narrative)
- why the user needs it (the goal)

12 Appendix 1 User Personas

12.1 Basque

Persona A

NAME	Josune
AGE	54
JOB ROLE	Vice minister of Public Health
LOCATION	Vitoria-Gazteiz, Basque Country, Spain
TRAITS	Negotiator, Communicative
BIO **GOALS** (objectives this person hopes to achieve through MIDAS)	Education: Business Studies. Master in Business Administration (MBA. Background experience in managing Public Health teams. She also has past experience in managing private company. She has a quite tight agenda of government meetings, health commissions and coordination meetings with province level public health epidemiologists. She has to manage many public health policies, so time and involvement on each topic is not what she would like to. She needs synthesised information to understand each prioritised health issues' status, intervention areas and policy status. She also need easy to convey synthesised reports, to built trust with society and other political parties. • To take data-grounded decisions, but having main intervention areas already synthesised • To reason the adopted data-grounded policy decision • To present data-grounded policy forecast • To present evolution of implemented policies
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS) **TECHNOLOGY SKILLS (% level)**	• Not up-to-date and population level data • Lack trust of in adopted policies • Lack of reasoning of adopteLack of simple and visual means to convey data in health comission and to society in generald policies ***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	45%
Software Mobile Apps Social Networks **DATA ANALYTICS SKILLS (% level)**	45% 45% 25% ***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	0%
Analytics	0%
Visualization	40%
Interpretation	50%

Persona B

NAME	Mikel
AGE	46
JOB ROLE	Province-level public health epidemiology manager
LOCATION	Bilbao, Basque Country, Spain
TRAITS	Clear talking, Fact-grounded talking
BIO	Education: Master degree on Epidemiology, Bachelor degree on Mathematics. Background experience in managing cancer and mortality registry, responsible for specific health issue managing within public health. Given the many possible areas of intervention of public health, actually he needs to find, get access, understand and cross different data sets before he starts interpreting the available data. Many of the data that he can access has been captured in limited surveys. He is willing to work with tools that would point him out candidate intervention areas and variables, and will allow him to forecast policy options and monitor implemented policy's evolution.
GOALS (objectives this person hopes to achieve through MIDAS)	• To have rate-adjusted detailed descriptives of health issues • To cluster population in intervention profiles • To identify intervention areas and variables • To forecast and evaluate policy outcome • To monitor implemented policies evolution
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	• Working with data limited to samples / survey-based • Diffulties to analyse heterogeneous data variables • Lack of support identifying candidate intervention areas / identifying new trends / patterns • Lack of tools to forecast / evaluate policy options
TECHNOLOGY SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	60%
Software	60%
Mobile Apps	60%
Social Networks	20%
DATA ANALYTICS SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	10%
Analytics	750%
Visualization	30%
Interpretation	100%

Grant Agreement No: 727721

Persona C

NAME	Lourdes
AGE	38
JOB ROLE	Pediatrician responsible of an assigned children patient
LOCATION	Bilbao, Basque Country, Spain
TRAITS	Devoted, kind
BIO	Education: Medical degree and a specialisation on pediartry. 10 years pediatrician consultant experience and has participated in heatlh research projects. She is concerned of classical activity and food recommendation's limited effect and the increase of overweight and obesity cases. Eventhough she doesn't know her group compares to other nearby groups or general Basque Country indicators. She's willing to actively contribute to intervention, but it's not clear to her where to intervene besides classical recommendations of which other factors to capture in consultation to try to contribute to define issue interventions.
GOALS (objectives this person hopes to achieve through MIDAS)	• To know her group overweight / incidence within region / province / country • To know which other variables she can capture for further analysis • To know which specific intervention areas to work in her group / context
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	• Lack of updated picture of her group state within a the basque context • Lack of group specific intervention areas
TECHNOLOGY SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	25%
Software	30%
Mobile Apps	40%
Social Networks	40%
DATA ANALYTICS SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	0%
Analytics	0%
Visualization	0%
Interpretation	50%

Grant Agreement No: 727721

12.2 Finland

Persona A

Name	Jenni	**BIO**	Jenni is the director of health services, with many years of experience in the field. She has a background as a doctor with basic training. Currently, she is responsible for organizing health services, mental health & substance abuse services in city level. Additionally, she participates as an expert in the national healthcare reform project from a regional level also. In personal life, Jenni has twin daughters (15 year olds) and an elder son (20 year old). She contemplates various necessities for mental health case comparing in her own experience and her kids teenage growing up.
Age	55		
Occupation	Director of health services, City of Oulu		
Status	Married		
Location	Oulu, Finland		
Tier			
Archetype			
Quote	We don't have information about cost and impact yet and what is the time window during which we can see some investment, it's always said that one euro invested in substance abuse services saves six euros but what is the time window there.		
MOTIVATIONS		**PERSONALITY (%)**	
Incentive		Extrovert – Introvert	Extrovert 80%
Fear		Sensing – Intuition	Sensing 75%
Achievement		Thinking – Feeling	Thinking 70% - feeling 30%
Growth	Yes (minor motivation)	Judging – Perceiving	Judging 80%
Power			
Social	Yes (major motivation)	**TECHNOLOGY (level)**	
		IT and Internet	60%
GOALS		Software	50%
a.	Organizing health services in city level (resource allocation)	Mobile Apps	50%
b.	Organizing mental health & substance abuse services in city level	Social Networks	40%
c.	Facilitate the national healthcare reform from regional level	**Data analytics skills**	
FRUSTRATIONS		Integration	30% (Yes, to some extent)
a.	Most of the relevant data is in silos, integration from multiple sources is a challenge.	Analytics	20% (understands the need, still lacks technical aspects)
b.	generalized preventive mental health services might hinder target group specification	visualization	20% (uses manual visualizations using general tools, eg. Excel, powerpoint, etc.)

Grant Agreement No: 727721

Persona B

Name	Pia Majala	**BIO**	Ms. Majala is a regional politician who is currently working as the Director of Education for Kuusamo region, Finland. She has been involved in public service for more than 30 years. Her work mainly focuses on organizing education related services, but at the same time she contributes to the physical and mental well-being of young people as a politician. She has academic degree specialising in Education. Also, she has worked in multiple projects focusing on wellbeing and mental health in her long public service career.
Age	58		
Occupation	Municipal Politician, Director of Education		
Status	Married		
Location	Kuusamo, Finland		
Tier			
Archetype			
Quote	I feel like utilising data, is still today, not systematic in municipal decision-making. It's pretty separate, so when I make preparations there are certain things of course, I utilise it for certain decision making situations. But, it's not systematic in any way.		
MOTIVATIONS		**PERSONALITY (%)**	
Incentive	minor	Extrovert – Introvert	Extrovert 70%
Fear		Sensing – Intuition	Intuition 40%
Achievement	minor	Thinking – Feeling	Thinking 50%
Growth		Judging – Perceiving	Perceiving 30%
Power	Major		
Social	Major	**TECHNOLOGY (level)**	
		IT and Internet	60%
GOALS		Software	50%
a.	Collect relevant data from multiple sources and make sense out them through comparative analysis.	Mobile Apps	40%
b.	To create a more data oriented decision making practice in the regional level.	Social Networks	40%
c.	To identify and utilize more variables that are relevant to youth wellbeing, besides education data and financial figures.	**Data analytics skills**	
FRUSTRATIONS		Integration	20% (manual integration)
a.	Lack of comprehensive dataset and up to date information on youth people.	Analytics	

midas
Meaningful Integration of Data, Analytics and Services

Grant Agreement No: 727721

Persona C

Name	Markku Järvinen	**BIO**	Mr. Järvinen is an elected Member of Finnish Parliament. He is affiliated with one of the major political parties, who is part of the government. As a seasoned politician, Markku has participated in different policy level projects for different domains. However, his involvement in the healthcare arena is rather recent. Currently he is involved in the national healthcare reforms that is ongoing in Finland. His academic background lies in political science. He understands the complexity and importance of whole healthcare sector and emphasize on the need to improve services. Personally, he is not so much tech savvy, for those purposes he depends on parliamentary data analysts and interprets on the analyses.
Age	48		
Occupation	Member of Parliament, Finland		
Status			
Location	Helsinki, Finland		
Tier			
Archetype			
Quote	Healthcare is one of the area that is very close to every individual of the population. For Finland, we need to keep the services world class and make sure our services are up to the task always.		
MOTIVATIONS		**PERSONALITY (%)**	
Incentive		Extrovert – Introvert	Extrovert 100%
Fear		Sensing – Intuition	Intuition 60%
Achievement	Yes	Thinking – Feeling	Feeling 50%
Growth	Yes	Judging – Perceiving	Perceiving 50%
Power	Yes		
Social	Yes	**TECHNOLOGY (level)**	
		IT and Internet	30%
GOALS		Software	20%
a.	Steer the healthcare reforms work and service integration from personal experience.	Mobile Apps	40%
b.	Familiarize and accustom himself with new phenomena like MyData, data analytics, data integration.	Social Networks	50%
c.	To lead and create a culture of using data integration and analytics in the higher level.	**Data analytics skills**	
FRUSTRATIONS		Integration	0

Persona D

Name	Hanna	**BIO**	Hanna is a researcher working at THL. She has spent all of her work career at THL and even did her practical training there, before graduation. Her work mainly focuses on scientific research on social group differences in health determinants such as education, income and health behaviour. She has academic degree specialising in Health Sciences. She is married and has two children.
Age	35		
Occupation	Researcher, THL		
Status	Married		
Location	Helsinki		
Tier			
Archetype			
Quote	Information on health and welfare goes beyond static figures and graphs. It is important to be able to examine self defined sets of variables, and make estimates, based on reliable, evidence based data.		
MOTIVATIONS		**PERSONALITY (%)**	
Incentive		Extrovert – Introvert	Introvert 70%
Fear		Sensing – Intuition	Sensing 80%
Achievement	Yes	Thinking – Feeling	Thinking 60%
Growth	Yes	Judging – Perceiving	Perceiving 60%
Power			
Social	Yes	**TECHNOLOGY (level)**	
		IT and Internet	70%
GOALS		Software	80%
a.	Spotting differences between social groups and addressing directives at them.	Mobile Apps	60%
b.	Trying to make the directives be included in policies.	Social Networks	90%
c.	Trying to make her message noticed on higher levels of administration.	**Data analytics skills**	
FRUSTRATIONS		Integration	70%

12.3 Northern Ireland

Persona A

NAME	Eilis McDaniel
AGE	54
JOB ROLE	Director, Family and Children's Policy Directorate, Department of Health
LOCATION	Belfast
TRAITS	Results-focussed; people-oriented.
BIO	Education: BA in Celtic Languages and Literature; MSc. in Computer Science and a Masters in Business Administration. Background: Worked in Irish Place-name research in Queen's University Belfast at the start of her career. Following that, she worked as a software developer for around 10 years before moving into policy development in central government. Within the Department of Health, she has lead responsibility for policy and legislation relating to family support, child protection, looked after children and adoption. She is also leading on the implementation of an Early Intervention Transformation Programme – a cross-departmental initiative to provide early support to families and children in need. Previously, she worked in the Children and Young People's Unit in the former Office of the First and deputy First Ministers, where she was responsible for driving forward the over-arching Northern Ireland 10-year Strategy for Children & Young People. She has also worked as Private Secretary to a former First Minister.
GOALS (objectives this person hopes to achieve through MIDAS)	The development of robust KPIs, improved measurement of outcomes; better designed services and provision; the right services in the right place; better informed policies and strategies to meet the needs of vulnerable children and young people, including those in care; benchmarking good practice across HSC services and HSC Trust areas and establishing consistency of approach; revised Delegated Statutory Functions reporting (i.e. more efficient monitoring and analysis).
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	DoH is restricted to 6 monthly and annual "snapshot" data (gathered at a single point in time) rather than longitudinal data. It is therefore difficult to establish patterns and historical trends with ease, making effective policy decision-making and responsive service planning and delivery more difficult to deliver.
TECHNOLOGY SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	70%
Software	50%
Mobile Apps	50%
Social Networks	50%
DATA ANALYTICS SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	40%
Analytics	40%
Visualization	40%
Interpretation	40%

Persona B

NAME	Susan Campbell
AGE	55
JOB ROLE	Head of Regional Data Warehouse, HSC NI
LOCATION	Belfast
TRAITS	Optimistic; collaborative
BIO	Education – BSc in Physics and Computer Science. She has worked in Health Service IT for over 30 years. She has worked in application development and support for most of that time, with a particular interest in making information from operational systems available to users for reporting and analysis. She manages a small team within the Business Services Organisation that maintains and enhances the HSC Regional Data Warehouse, using a range of technologies. She also has responsibility for Integration services and the development of small-scale bespoke regional web applications for the HSC.
GOALS (objectives this person hopes to achieve through MIDAS)	The Regional Data Warehouse is mainly used by trained Information staff. Susan would like to make the data accessible to decision makers in the course of their day-to-day work.
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	Buying in to expensive proprietary toolsets which require high level of training, knowledge and expertise.
TECHNOLOGY SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	80%
Software	80%
Mobile Apps	50%
Social Networks	30%
DATA ANALYTICS SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	60%
Analytics	40%
Visualization	40%
Interpretation	20%

Grant Agreement No: 727721

Persona C

NAME	Heidi Rodgers
AGE	44
JOB ROLE	Deputy Principal Statistician
LOCATION	Department of Health, Belfast, Northern Ireland
TRAITS	Creative and supportive
BIO	Master in Psychology from the Norwegian University of Science and Technology Worked as a statistician for Department of Health the last 8 years, focussing on children's social care the last 5 years. Enjoys a 'hands-on' approach when working with datasets, to be able to see trends that cannot be picked up from running analysis or a report.
GOALS (objectives this person hopes to achieve through MIDAS)	Establish links between datasets that can set precedence for future work.
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	There is no fast-track, it always takes longer than expected to go from A to B, B to C, C to…
TECHNOLOGY SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	70%
Software	70%
Mobile Apps	50%
Social Networks	50%
DATA ANALYTICS SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	70%
Analytics	80%
Visualization	80%
Interpretation	80%

Grant Agreement No: 727721

12.4 Republic of Ireland

Persona A

NAME	Kathleen
AGE	49
JOB ROLE	Clinical Nurse Specialist, Health Service Executive (HSE)
LOCATION	Letterkenny, Co. Donegal, Ireland
TRAITS	Clinical decision maker, educator
BIO	Kathleen is a qualified nurse who specialises in Diabetes. She provides care for persons with diabetes in the community (80%) and in the acute hospital environment (20%). Kathleen has done a Degree in Nursing Studies and a Higher Diploma in Diabetes Management as well as a Postgraduate Diploma in Nurse Prescribing of Medicinal Products. Kathleen needs to be able to identify the cohort of patients with diabetes in her geographical area. She needs to be able to access records on their current and past episodes of care to determine health needs for now and the future. Kathleen needs to put in place a plan of care for each person which is determined by an assessment and a priority rating depending on the risk to health. Kathleen needs to be reassured that the person is managing and supported to self care and receives guidance in medication administration where needed, has access to education to manage possible fluctuations in condition such as hypo or hyperglycaemic episodes.
GOALS (objectives this person hopes to achieve through MIDAS) **FRUSTRATIONS** (the pain points they'd like to avoid through MIDAS) **TECHNOLOGY SKILLS (% level)**	• To have access to a national diabetes register with up to date information on all persons with diabetes in her region • To support self care through sharing information relating to diabetes care in real time providing supported decision making for the person with diabetes • Lack of access to health records at the point of care or lack of sharing of information across hospital/community boundary • Lack of IT software to facilitate information recording in a health record which is accessible for others • Lack of database to support caseload management (Prioritisation by risk and health needs) • Statistical and trend analysis of care episodes to determine risks ***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	70%
Software	40% (Silo systems)
Mobile Apps	20% (Not always feasible on HSE Systems)
Social Networks	30% (To support patient engagement)
DATA ANALYTICS SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	20% (As works with silo systems, this can prove difficult)
Analytics	10% (Lacks technical aspect, but regularly analyses data manually)
Visualization	30% (Has ability to use graphs in Excel and Powerpoint to show data)
Interpretation	30% (Lacks detailed knowledge on measurement and graph types to show data in different formats)

Grant Agreement No: 727721

Persona B

NAME	Mary
AGE	34
JOB ROLE	Practice Development
LOCATION	University Hospital, Ireland
TRAITS	Resilience
BIO	Mary is a qualified nurse with a Degree in Quality Management. She is focused on having access to up to date information that provides a picture of the quality of care in her hospital. Mary has to assure that policies and standards are being maintained in delivering quality care. She needs access to accurate information on key performance indicators relating to priority risk areas in her hospital such as: the number of adverse events causing injury or death, number of hospital acquired infections or pressure ulcers, number of admissions per day and percentage bed occupancy, number of complaints, staffing levels and acuity levels of patients for that day.
GOALS (objectives this person hopes to achieve through MIDAS)	Needs information from a range of sources amalgamated in a single workspace or dashboard to support decision making Needs quality information which is evidenced based and supported by quality standards which gives her valuable information to determine the level of risk on a ward or in the hospital on any given day or period
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	Access to standardised quality data, following set criteria for collection and validation Real time data which can be comparatively analysed and utilised and trends identified
TECHNOLOGY SKILLS (% level) IT and Internet Software Mobile Apps Social Networks **DATA ANALYTICS SKILLS (% level)**	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)*** 80% 40% 20% 10% ***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	40%
Analytics	20%
Visualization	20%
Interpretation	20%

Persona C

NAME	Dr Conor O' Shea
AGE	58
JOB ROLE	General Practitioner (GP)
LOCATION	Wheaton Hall Medical Practice, Drogheda, Dublin, Ireland
TRAITS	Communicator,
BIO	Education: Medical Degree, Member Irish College of General Practitioners (ICGP) and Fellow of Royal College of General Practitioners (RCGP). The Medical Practice would have a population of 8000 with 10% in the over 65 age bracket. The focus for the practice would be on general population health with specialist clinics for diabetes, asthma, dermatology, blood pressure and cardiac monitoring and general health screening. As a GP, he has access to internal medical records for patients registered with the practice, medication listings and laboratory reports for tests ordered from the practice. Wider integrated access to summary information would be ideal for diagnostic or planning care such as in patient episodes of care in acute hospitals.
GOALS (objectives this person hopes to achieve through MIDAS)	• Optimise patient care through having the right information available at the time of the visit and access to appropriate information to support diagnosis and treatment • Improve communication channels for all interested parties in all directions • Improve access to information for the patient to support self care
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	• Lack of Integration of data across healthcare providers • Require greater shared policies for care following patient journey • Need resources to support and involve patient in own care
TECHNOLOGY SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	80% (Good access and skill)
Software	80% (Good access to GP Software)
Mobile Apps	20% (Not routinely used but used when appropriate)
Social Networks	60% (Strong part of medical practice role)
DATA ANALYTICS SKILLS (% level) Integration	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)*** 70% (Of data available in GP System, good integration but would benefit from external sources of information integration)
Analytics	70% (GP System supports good use of analytics)
Visualization	50% (System would benefit from more modern approach to visual display
Interpretation	70% (Supports decision making and interpretation)

12.5 Co-Design Workshop Personas

Persona A

NAME	Sean O'Mahoney
AGE	50
JOB ROLE	Finance / Resource Management
LOCATION	Dublin, Ireland
TRAITS	Accountant, Strategist
BIO	Education: Financial management. Some experience working in healthcare, local healthcare manager. Moved to Nationwide finance controller position with responsibility for government resource allocation / management.
GOALS (objectives this person hopes to achieve through MIDAS)	· Allocate well resources · Maximise value for available budget · Better population health outcomes for the available budget · Workforce distribution – balance
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	· Never enough money · Misallocation of resources · Lack of timely data for decision making · Lack of prioritisation · People in some areas complaining about long queues · Paper-based information to make decisions

Persona B

NAME	Julia
AGE	50
JOB ROLE	CEO of charity
LOCATION	
TRAITS	Communicative and negotiator
BIO	MBA or similar and passionate about the topic of the charity
GOALS (objectives this person hopes to achieve through MIDAS)	Consumer of information to raise information. Commision research and get evidence for fundraising. Educational perspective for the member. Help inference policy. Advantage of being closer to population, easier to gather personal data.
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	Not availability for integrated data. Sharing information among charities.
TECHNOLOGY SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
IT and Internet	60%
Software	40%
Mobile Apps	60%
Social Networks	80%
DATA ANALYTICS SKILLS (% level)	***(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)***
Integration	40%
Analytics	40%
Visualization	40%
Interpretation	40%

Persona C

NAME	James
AGE	30s
JOB ROLE	Previously large industry, now unemployed
LOCATION	
TRAITS	Has current conditions and thus health interest heightened (eg COPD)
BIO	May get access via GP May want printouts Might not have personal access but would want the detail, the impact, the purpose.
GOALS (objectives this person hopes to achieve through MIDAS)	· Patient choice? Who should they be able to review their doc or specialist – · Wants to change life and get better, healthier, job · Want to understand their conditions and what they can do about it · Could benefit be that he can see impact of interventions? · Could it be seeing the positive benefit or impact of policies · Self care and looking after own condition
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	· Can't choose doctors · Don't know if they are good

Persona D

NAME	Alex
AGE	28
JOB ROLE	Cyber-security expert
LOCATION	Belfast
TRAITS	
BIO	Tech savvy – grown up in mobile age with social networks. Knowledgeable about their own data and risks of sharing. Works in tech industry and early adopter of new technology. MyData enthusiast, want to control their own data and their own healthcare.
GOALS (objectives this person hopes to achieve through MIDAS)	Comparing their own data with population-wide statistics to see how they rank. Want to play with the data themselves so that they can understand policy decisions – can use as part of engagement process and help to change their minds. Want to have better control over their own healthcare and their data.
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	Communication – not understanding language or not getting results/information back from clinical settings Currently their personal data is being commercially exploited Don't understand why policy decisions are made
TECHNOLOGY SKILLS (% level)	(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)
IT and Internet	90%
Software	90%
Mobile Apps	80%
Social Networks	80%
DATA ANALYTICS SKILLS (% level)	(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)
Integration	20%
Analytics	30%
Visualization	50%
Interpretation	50%

Persona E

NAME	
AGE	
JOB ROLE	Director of Well-Being
LOCATION	Small town in rural area, reports to town mayor
TRAITS	
BIO	Social sciences, Humanities, Psychology
GOALS (objectives this person hopes to achieve through MIDAS)	● Integrate data from different sectors and service units (green – basic city services, yellow – specific needs, red – health care issues), ● predict the trends, ● to have a digital system with analytics and visualization for easy use
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	● General health of people in the city is decreasing ● Lack of available data integration ● Lack of communication ● Time consuming manual data work
TECHNOLOGY SKILLS (% level)	(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)
IT and Internet	50%
Software	50%
Mobile Apps	40%
Social Networks	40%
DATA ANALYTICS SKILLS (% level)	(some additional specifics here for each aspect as per Finnish colleagues completed User Personas would be beneficial)
Integration	40%
Analytics	30%
Visualization	30%
Interpretation	50%

13 Appendix 2 Requirements for custom widget

Inputs:

ID of a Div DOM element: Render any custom visualization within this container Div. You can use any technology (HTML Canvas, SVG, D3 etc.) to render the visualization unless HTML5 and all the major browsers support it.

Data Model: the data model for rendering the visualization. Currently, a generic data model is used to render all the visualizations in the MIDAS dashboard. If needed, standardized data-model specifications such as Vega-lite, can be used to define the data model.

Plugin should be called easily. For example:

```
customPlugin.renderVisualization("graphDiv1", dataModel);
```

Output:

The plugin will be responsible to render the visualization within the div (input div) and to provide all the required interactions (zoom, pan, data point selection etc.). Keep in mind that dashboard widgets are resizable so that the plugin should handle resizing of the visualization according to the new size of the container Div.

Data Model

```
var dataModel = {
        chartStaticImgUrl: "",
        chartTitle: "",
        chartSubTitle: "",
        xLable: "",
        yLable: "",
        valueRange: null,
        timeUnit: null,
        startDate: null,
        endDate: null,
        chartData: {
           xAxisType: "",
           yAxisType: "",
           xData: [],
           yData: []
        }
    };
```

Description of data model fields:

Field	Data Type	Definition
chartStaticImgUrl	string	url of the generated static image
chartTitle	string	Main title of the graph
chartSubTitle	string	string of object of interest names separated by comma
xLable	string	xTitle
yLable	string	yTitle
valueRange	JSON Array	Contains min and max values for Y-axis if required. for example [0, 100]
timeUnit	string	one of: “sec, “min”,”hour”,”day”,year (if required)
startDate	DateTime	Date/time in format: 2017-03-17 09:04:28.0 (if required)
endDate	DateTime	Date/time in format: 2017-03-17 09:04:28.0 (if required)
chartData	JSON Object	Contains all the required data to draw a graph. Explanation below.

chartData:

Field	Data Type	Definition
xAxisType	string	To calculate X-axis scale so that functionalities like zooming, panning and Inter-graph synchronization can be achieved. One of: integer, float, string, datetime.
yAxisType	string	To correctly calculate Y-axis scale so that functionalities like zooming, panning and Inter-graph synchronization can be achieved. type: integer or float
xData	JSON Array	Array of values for X-Axis
yData	JSON Array	It is an array of arrays. Explanation below

yData:

Field	Data Type	Definition
yData [0]	JSON Array of Objects	Array on zero index defines the type of visualization to be drawn on Y-axis. for example: [{ "seriesType":"line",label":"%"} { "seriesType":"line", ":"mean"}] SeriesType: defines the type of particular series. one of the following values: line, scatter, bar, horizontalbar, stackedbar, horizontalstackedbar, pie, radar, bubble, candle, area, timeline, correlationmatrix. label: defines a label for series used for the legend and

		data point tooltips.
yData[1…]	JSON Array of series data	Each subsequent array in the yData array, represents data of defined series. So according to above example first array (yData[1]) will contain data for “%” series and second array (yData[2]) contains data for “mean” series. Make sure the order of data in these arrays should be according to the order of data in xData array.

Visualization support with current data model:

The above data model supports following visualizations. It also supports combining of multiple visualizations into one graph widget. For example, combine line and scatter plot visualizations in one graph object.

1. Line graph
2. Scatterplot
3. Bar chart
4. Horizontal bar chart,
5. Stacked bar chart
6. Horizontal stacked bar chart
7. Pie chart
8. Radar chart
9. Bubble chart
10. Candle chart
11. Area chart
12. Timeline
13. Correlation matrix.

14 Appendix 3 Conceptual OpenVA Data Model

Figure A2.1 shows the conceptual data model of OpenVA. The data model is capable to handle both data with oi_measuredproperty* components and metadata with the rest of the components shown at Metadata layer. Metadata components are independent of the data components and can be also utilized without them. Dashboard layer shows utilization of Metadata layer from the User Interface (UI) point of view. The data model enables transforming background and time series data of virtually any kind into unambiguous structure to make a clear interface between data and the rest of the system.

Grant Agreement No: 727721

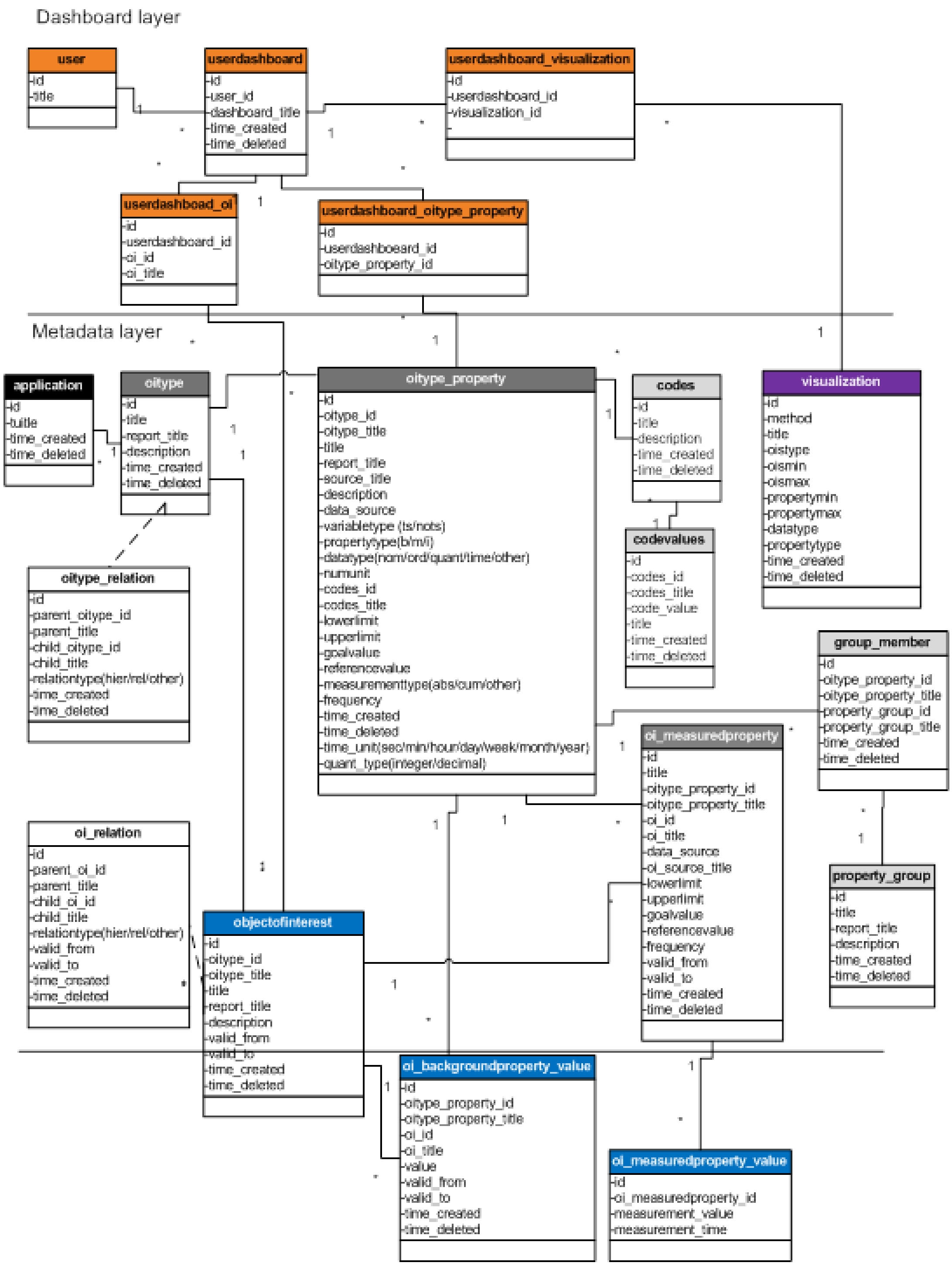


Figure A2.1: Conceptual data model of OpenVA