Ref. Ares(2015)1151516 - 16/03/2015

DELIVERABLE 1.1

State of the art Report

PROJECT NUMBER: 619186
START DATE OF PROJECT: 01/03/2014
DURATION: 42 months

DAIAD is a research project funded by European Commission's 7th Framework Programme.

The information in this document reflects the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.

Dissemination Level	Public
Due Date of Deliverable	Month 4, 30/06/2014
Actual Submission Date	22/07/2014
Work Package	WP1 'DAIAD Requirements, Architecture and Integration'
Task	T 1.1
Type	Report
Approval Status	Submitted for approval
Version	1.0
Number of Pages	130
Filename	D1.1_State_of_the_art_Report.pdf

Abstract

This deliverable provides a comprehensive review of state of the art regarding all aspects of real-time water monitoring (sensors, interfaces, data management, knowledge extraction, water demand management, pricing policies, etc.). It presents the key concepts, standards, approaches and tools on the respective areas and serves as a common knowledge based for research and development work in the project.

History

version	date	reason	revised by
0.1	14/5/2014	First version and section outline	Spiros Athanasiou
0.2	18/6/2014	Updates in Section 3, theme improvements	Spiros Athanasiou
0.3	28/6/2014	Updates in Section 3	Giorgos Giannopoulos
0.4	14/7/2014	Updates in Section 4	Christian Sartorius, Anja Peters
0.5	15/7/2014	Updates in Section 1	Thorsten Staake, Thomas Stiefmeier, Anna Kupfer
0.6	15/7/2014	Updates in Section 3	Yannis Kouvaras, Giorgos Giannopoulos
0.7	16/7/2014	Updates in Section 4	Lydia Dant
0.8	18/7/2014	Updates in Section 1	Anna Kupfer
0.9	21/7/2014	Updates in Section 2	Anna Kupfer, Thorsten Staake
1.0	22/7/2014	Various edits, fixed references, and introductions	Spiros Athanasiou, Giorgos Giannopoulos

Author list

organization	name	contact information
ATHENA	Spiros Athanasiou	spathan@imis.athena-innovation.gr
ATHENA	Giorgos Giannopoulos	giann@imis.athena-innovation.gr
ATHENA	Yannis Kouvaras	jkouvar@imis.athena-innovation.gr
BAMBERG	Thorsten Staake	anna.kupfer@uni-bamberg.de
BAMBERG	Anna Kupfer	thorsten.staake@uni-bamberg.de
AMPHIRO	Thomas Stiefmeier	stiefmeier@amphiro.com
ISI	Christian Sartorius	Christian.Sartorius@isi.fraunhofer.de
ISI	Anja Peters	Anja.Peters@isi.fraunhofer.de
WATERWISE	Lydia Dant	LDant@waterwise.org.uk

Executive Summary

This document reviews the state of the art in real-time water monitoring, across all involved scientific domains and disciplines. Our focus is on monitoring *domestic water consumption*, and our presentation follows the *lifecycle of water data*. We explore the various stages where water consumption is measured, managed, analyzed, and converted to knowledge, highlighting the established technologies and systematic approaches. In this respect we present the established practices of the Water domain, and also introduce concepts and paradigms from the ICT sector which will be applied in DAIAD under a novel concept.

As such, this report has been prepared with a dual purpose. First, serve a common knowledge base for the project, combining expertise, know-how and excellence from the Consortium in their respective domains. Second, serve as a starting point for researchers from the Water and ICT domains, by introducing relevant theory and technologies from these disciplines, highlighting potential common research areas and opportunities for closer collaboration.

The layout of the document is the following.

In Chapter 1, we present the state-of-the-art in water sensing technologies and practices. We first provide background information on conventional water meters, and then present in more detail technologies for single-fixture and multiple-fixture water sensing in domestic environments. We follow with a discussion of approaches for power supply in water sensing arrays, and the various connectivity options for transmitting water consumption data across the literature. Finally, a brief overview of a technologically and market-related field, the Smart Home, is provided.

In Chapter 2, we present the state-of-the-art in the novel area of interventions for water consumption. First we highlight specific literature and insights from behavioral psychology regarding feedback and stimuli. In the following, we provide specific empirical evidence for interventions to convey water consumption, districting them in the household and fixture levels.

In Chapter 3, we introduce water consumption data in the Big Data landscape. First, a discussion on the characteristics and challenges of Big Data is provided. In the following, we delve into two major data management paradigms applied in the Big Data domain, NoSQL and the MapReduce framework. Finally, we elaborate on various data analysis approaches which can be applied in real-time water consumption data, with an emphasis on recommendations and personalization.

In Chapter 4, we focus on Water Demand Management, presenting both the current state-of-the-art in the literature, in addition to highlighting challenges that can be addressed from the project. First, we discuss the determinants of water demand as identified in the literature and practice. In the following, we revisit interventions for altering water demand, focusing on structural and psychological determinants. Finally, we provide an overview of the most recent and relevant large-scale trials of smart water metering technologies, providing and discussing insights gained.

Abbreviations and Acronyms

AMR	Advanced Metering
API	Application Programming Interface
BAN	Building Area Network
DAG	Directed Acyclic Graph
DBMS	DataBase Management System
DSL	Digital Subscriber Lines
GPRS	General Packet Radio Service
GSM	Global System for Mobile
HAN	Home Area Network
HDFS	Hadoop File System
ICT	Information and Communication Technologies
IoT	Internet of Things
JSON	JavaScript Object Notation
LED	Light Emitting Diode
MEMS	Micro-Electro-Mechanical Systems
NAN	Neighborhood Area Network
PAN	Personal Area Network
PLC	Power Line Communication
RDBMS	Relational DataBase Management System
RDF	Resource Description Framework
REST	Representational State Transfer
SQL	Structured Query Language
SVM	State Vector Machine
WLAN	Wireless Local Area Network
XML	Extensible Markup Language
YARN	Yet Another Resource Negotiator

Table of Contents

1. Water Sensing Technologies

Mechanisms to preserve water by means of mechanical flow and volume restrictors are currently integrated in virtually every sanitary product, including low-flow faucets, eco-showerheads, and two-mode flushes. They are promoted even in countries where water is abundant, since many derivates also conserve *hot water* and thus *heat energy*. While their adoption is vital to water and energy efficient societies, their potential to trigger further savings is limited, as the flow rates of the existing eco-products approached a minimum acceptable by customers.

In order to achieve the necessary savings that go beyond the effects from state-of-the-art mechanical systems, high hopes are being placed on ICT-based systems that can *measure, communicate, and display consumption information in real-time and for individual water outlets*. *Data-driven* approaches can motivate consumers to reduce the time over which water is extracted from the network. Moreover, the provided information can help to increase awareness among citizens by showing them that *behavior* has an immediate impact on overall demand. Feedback interventions targeting water usage are even expected to have a considerably stronger effect on behavior than related approaches for electricity given the suitability of water consumption as target variable: Curtailment behavior is more simple for water than it is for electricity (faucet closed = no water consumption vs. many consuming electrical devices), water consumption is easier to attribute to the user (individual extractions from faucets vs. high baseline consumption by electrical appliances), and the units of measurement are more tangible (liters vs. kWh). In fact, first lab trials and small-scale field studies have shown very strong effects from water consumption feedback with savings above 20% on top of established mechanical flow restrictors as long as user interfaces apply suitable psychological mechanisms [SF11].

The realization of adequate feedback interventions and water management tools depends on the granularity of the available consumption data. Currently, the installed water meters on a household level provide very limited data for conservation campaigns and agile water management system. In this context, a number of sensors and sensor arrays have been applied for *water measurement in domestic environments*. Recent advances in micro-electro-mechanical systems (*mems*) have enabled the miniaturization and innovative packaging of water flow sensors (e.g., in the form of thermal mass flow meters). In addition, non-traditional sensing technologies (e.g., using ambient noise measurement) have been outlined in the literature. These approaches are being combined with data transmission techniques of any sort. The common goal is to provide at the least possible cost (also in terms of energy consumption and installation complexity) systems with suitable measurement precision for volume and flow rate (and often also temperature if energy consumption is concerned).

For the further discussion, we can categorize existing work on measuring water consumption in domestic environments in two areas: (a) water sensing for a *single fixture* (e.g., on the faucet-level), and (b) sensing water consumption for a *whole residence*. Research in the field of residential sensing currently focuses on sensing and monitoring water consumption at the household level, with the aim of providing *accurate water measurements per mains inlet.* In order to provide feedback on the level of individual faucets, a number of approaches are documented in the literature, with varying levels of success and complexity. At the one side of the spectrum, approaches are focused on extracting information from a single sensing point through data disaggregation (*centralized systems*). At the other side, a (possible large) number of additional sensors at individual fixtures are used (*distributed sensing*), with the solution resembling a networked system of sensors that are able to communicate.

With the DAIAD vision in mind, this chapter starts with outlining conventional metering technologies. Due to their intrusiveness and often-limited connectivity, further research on domestic sensing technologies is explored. This includes the potential of conventional methods and centralized approaches as well as novel techniques with the goal to obtain highly granular data on individual extraction points. Therefore, technologies on fixture level are investigated as well. As one major aspect of DAIAD concerns the long-term application of ICT-based concepts in the field, an overview of power supply for sensing and communication modules is provided as well. DAIAD aims to integrate sensors with feedback mechanisms in order to enhance customer and utility action. Thus, technologies that ensure an interaction between sensors and data collectors are included, and the chapter closes with integration-aspects of water metering, smart home, and home automation concepts.

1.1. Conventional water meters

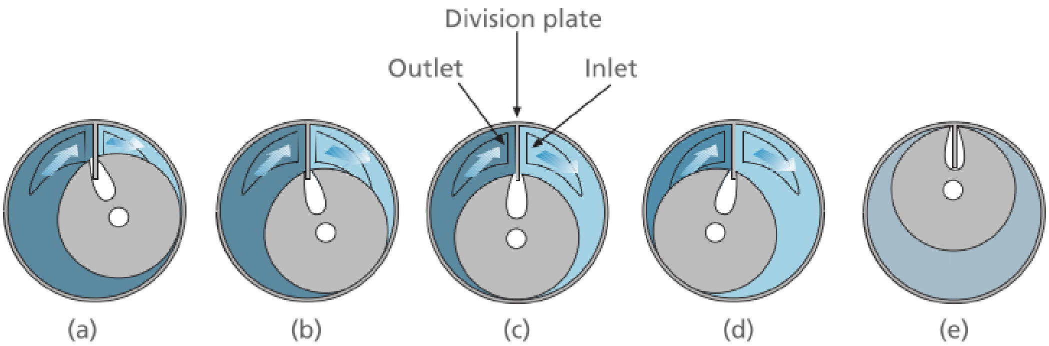
Traditionally, the goal of domestic water sensing infrastructures has been to support billing processes of utility companies and landlords by providing information on the amount of consumed water per residence. Consequently, conventional water meters only gathered aggregated data on the total per residency. The provided data only allows for visualizing consolidated consumption information to residents for all domestic activities in one bill with frequency as low as one per year. When confronted with the total water consumption of an entire year, typically neither can consumers establish a relation between a specific behavior and their impact on water demand, nor can they derive measures to cut their consumption. They are also not enabled to differentiate seasonal or fixture based water demand. In order to improve the consumption feedback, it is necessary to provide data more frequently. As many basic principles to provide such information are available, this section gives an overview on main sensing mechanisms, including mechanical, electromagnetic, and ultrasonic meters, along with specific advantages and disadvantage of each technique.

1.1.1. Mechanical water meters

Mechanical meters include volumetric, inferential, and combination meters. The domestic domain primarily deploys mechanical meters, whereas large water consumers, e.g. in the industry sector, typically install electromagnetic and ultrasonic water meters for wider pipe diameters and special accuracy needs [Zy11].

1.1.1.1. Volumetric water meters

Volumetric meters – also called positive displacement meters – measure volume by operating pistons and nutating discs [Zy11] [SB04]. They measure the water flow that is running through a fixed compartment with a known capacity of water volume. Pistons or nutating discs control the inlet and outlet of the measuring chamber by acting as valve and displacement devices. Each cycle allows a specific amount of water to flow through the compartment. Picture 1 illustrates the operating principal in a simplified way. Due to the water flow, the disc is moved clockwise in the measurement chamber. It has the size to cover the inlet and the outlet, therefore only the desired amount of water is stuck in the compartment (e). When water is being transported through the pipe, the disc is pushed to open the inlet and rotates the disc (a). The incoming water is separated from the remaining by the division plate. The increasing amount of water forces the disc to uncover the inlet gradually (b). Subsequent to the inflow of water, the disc starts to cover the outlet and the remaining water continues to flow out of the water meter (d). When the disc moves to position (e), the whole chamber is filled again and the rotation of the disc is registered as an electrical signal and can be converted to the volume of water. The measurement results are generally highly accurate when the meter is used in an infrastructure with high water quality. Then the meter attests long lifetime at moderate costs. However, air and solids in the water endanger the high measuring accuracy and the durability of volumetric meters, respectively. For example, sand can get stuck in the meter blocking the disc rotation. The meter is mainly used for domestic water metering because it is suitable for low flows [Zy11][SB04]. With this measurement principle, it is possible to identify nuances of flow differences, which makes it an interesting candidate for identifying single extraction processes and even for disaggregating individual appliances or users actions from a central point sensor at the mains inlet. Higher pressure drops at high flow rates (e.g., from parallel extraction events) are a downside of the volumetric approach.



Picture 1: Illustration of one piston rotation phase for volumetric measuring [Zy11]

1.1.1.2. Inferential water meters

Inferential meters use water velocity as the central measuring unit for flow sensing. Inferential meters are typically categorized as:

- Single-jet meters,
- Multi-jet meters, and
- *Woltmann* meters [Zy11][SB04].

All three types of meters use the "watermill" mechanism of an impeller that measures the *velocity*. Single or multi-jet meters employ an impeller with radial vanes whereas the *Woltmann* meter uses helical vanes. The impeller spins with a certain speed depending on the water flow. The meter registers the velocity of the vane and infers the volumetric flow rate of the water. Domestic water sensing utilizes single/multi-jet meters, whereas industrial or large water consumers install *Woltmann* meters that are more robust.

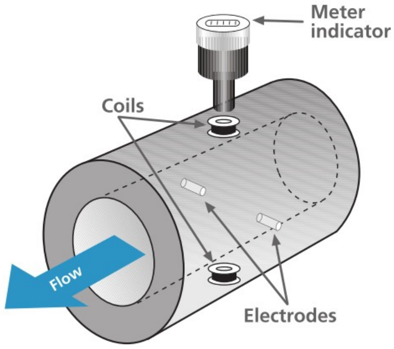
Accuracy of single-jet meters is susceptible to wear of the impeller during lifetime. Additionally, low flow rates and air passing through the impeller leads to under- and over-registrations. Also, reverse flows entail inaccuracy. Multiple-jet meters represent and improved approach to single-jet meters. The utilization of multiple jets drives the impeller in a more balanced way and thereby avoids one-sided wear. The more particles are in the water and the higher the velocity of the water is, the more problematic become wear considerations. In contrast to single-jet meters, multi-jet meters are also more tolerant to the pipe's velocity profile and to low flow rates. However, blocked jets fundamentally deteriorate the accuracy of multi-jet meters. Accuracy of *Woltmann* meters is only impeded by low flows – it is more robust against air bubbles and particles. However, both types of inferential meters are susceptible against flow disturbances that often occur as a pipe bends or after valves close to the meter section. Moreover, low flow rates are often not detected, which makes it more difficult to obtain high-resolution data in an important operating point of a smart sensing system. Yet inferential meters dominate the market for residential applications and thus are of major importance for this project.

1.1.1.3. Combination water meters

In order to improve the accuracy, a combination of an inferential and volumetric meter is associated with a valve [Zy11][SB04]. A multi-jet or Woltmann meter measures high flow rates, whereas a bypass meter that consists of a rotary piston or single/multi-jet meter handles low flow rates. The valve controls the entry to the inferential or the volumetric meter depending on the flow rate. The mechanisms allow a wide dynamic range, combining high accuracy at low flow rates with low pressure drops at high flow rates. Especially housing complexes, hospitals, and schools require such meters for a wide range of water flows.

1.1.2. Electromagnetic water meters

Electromagnetic meters build on *Faraday's Induction* law for measuring the velocity of water flow [Zy11][SB04]. In an electromagnetic meter, a magnetic field is applied to a tube, which causes a voltage difference that is proportional to the flow velocity perpendicular to the flux. Consequently, the metering principle is only applicable to conducting fluids, yet it works for water that contains ions. The inner pipe surface has to be non-conducting in order not to shortcut the induced voltage. Often, rubber-lined steel tubes are used. Coils induce a magnetic field at the top and the bottom of the pipe (Picture 2). Two electrodes at the right angle measure the induced voltage. If the polarity of the magnetic field remains constant, the electrodes deteriorate. Thus, permanent magnets are not suitable for this application. Alternating currents are needed to mitigate the problem by generating alternating fields; this increases the lifetime of the meters but considerably increases their power consumption.



Picture 2: Electromagnetic flow meter principle

Electromagnetic meters represent a non-intrusive measuring method for water flow. There does not exist any wear to the measuring parts due to the flow itself. Additionally, there is no pressure loss caused by the moving parts. Density, viscosity, temperature, and pressure do not limit the measurement accuracy. Alternating currents eliminate disturbance voltages. However, deposits (e.g., sand or chalk) may affect the sensing activity of the electrodes. Furthermore, air in the liquid and turbulences reduce the overall accuracy.

1.1.3. Ultrasonic water meters

Ultrasonic flow meters, categorized into *transit-time* and *Doppler* meters, sense the ultrasound signal sent between two transducers [Zy11][SB04]. For measuring water consumption, transit-time meters consist of two transducers installed with a certain distance between each other on the pipe. For the actual measurement, each transducer transmits an ultrasound signal to the other transducer. Depending on the flow direction of the water, the sound waves slow down or speed up in the measurement media. By means of registering the transit time of both transducers, the flow velocity can be inferred from the two measured time spans. The *Doppler* effect describes the frequency change of a sound wave when it is reflected back from a moving object such as sand or dirt in the water. First, frequency is sensed by a receiver and, secondly, related to velocity. Both meters are very cost effective for very large pipe diameters.

Accuracy of transit-time meters increases with the size of the pipe. With longer path lengths, the speed of the signal increases and the measurement of the flow rate becomes more accurate. However, such meters are sensitive to the velocity profile in the pipe, which means that steps must be taken to reduce flow turbulences. Besides, clean water is required to obtain reliable results. In contrast, the *Doppler* meter *needs* air bubbles or solids in the water flow for the measuring process. However, unequal flowing speed of the water and uneven distributions of particles limit accuracy. *Doppler* meters are mainly used for cases where billing is not applied but only flow monitoring is essential.

1.2. Single fixture sensing (decentralized/distributed systems)

Distributed solutions measure water consumption for single residential fixtures, such as bathroom faucets, and are typically tightly integrated with specific-purpose feedback mechanisms. They are often designed to provide instantaneous feedback that aims to motivate curtailment behavior among their users. However, many of the existing decentralized approaches lack interfaces to other systems due to restrictions in power supply. Thus, they do not provide consumption data to other system components nor enable data integration in a larger context. The subsequent section outlines fixture-level sensing techniques for water consumption that offer the capability to capture complex consumption patterns. Research prototypes and commercial devices mainly concentrate the shower as a main consumer of both water and energy, thus many examples come from this domain.

1.2.1. Faucet

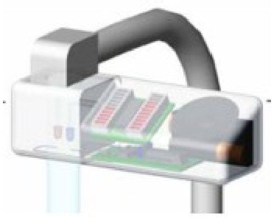
A basic approach for extraction-point sensing involves identifying the *operation* of domestic water fixtures to deduce water flow. For example, in most toilets a simple *switch* would be sufficient to approximate water demand since the amount of water consumed per usage can be known. More sophisticated systems measure and display *variable* outlets (i.e., regular faucets). Examples include *mechanical flow meters* (WaterBot [AB+05]) and *microphones* (UpStream [KP10]). Arroyo *et al*. developed a device called WaterBot that presents immediate feedback in the form of visual and auditory reminders (Picture 7). The device is to be installed on faucets to motivate people to turn off the tap when water is not used. They also yield the idea to sense water with a small turbine that also powers the display. However, no information could be found for realizing that idea. Kuzentsov & Pavlos proposed UpStream [KP10], a system combining low-cost water flow sensing coupled with persuasive displays (Picture 9). A microphone connected to an *Arduino* micro-controller measures water flow.

Picture 3: Collection bag for consumption measuring

Picture 4: Efergy shower timer

Picture 5: Ecosavers shower timer

Picture 6: Waterpebble shower timer

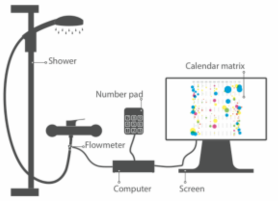


Picture 7: Waterbot installation

Picture 8: Show-me display

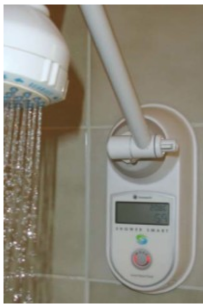
Picture 9: UpStream displays

Picture 10: WaiTEK shower monitor



Picture 11: The Shower Calendar

Picture 12: Amphiro shower display



Picture 13: Invetec shower meter

Picture 14: Amphiro a1

1.2.2. Shower

Available consumer products[1] for fixture-level sensing focus on the shower as point of measurement. They consist of devices measuring the time spent in or the water volume consumed during a shower [WS+10]. Simple shower timers such as the *ecosavers* (Picture 5) only measure the duration of a shower by the help of user entries (pushing a button at the beginning and at the end of a shower). In addition, an alarm tone can be set for exceeding a certain time. Other devices are able to infer the amount of water that is consumed. Using mostly low-cost, low-precision measurement modules – or even not measuring volume at all but relying on a constant flow rate and a timer – many devices require calibration process that are to be performed by the consumer. These devices are typically set up with the help of a collection bag with a known volume (Picture 3). One example of such a device is the shower monitor *efergy* (Picture 4). It displays the amount of water consumed during a shower, and an alarm can be set that is activated after exceeding a certain consumption limit. A progress bar shows the current state of consumption in relation to the desired limit. *Waterpebble* (Picture 6) senses water consumption at the drainage. A calibration process during the first shower enables a “traffic light” and an audible feedback. The green light phase – indicating

[1] http://www.amazon.de

good behavior – will be continuously reduced for subsequent showers in order to motivate gradually shorter showers. *Waterpebble* is an example of a low-tech, low price product that might replace simple shower timers.

In contrast to these end-user ready devices, the research community mostly concentrates on automatic consumption sensing without user input in order to increase accuracy. The literature focuses on direct sensing capabilities integrated into the shower installation. Various devices and measurement technologies can be identified:

- The WaiTEK shower monitor (Picture 10): It includes sensors to measure the water flow and the temperature. Long showers result into acoustic signals that motivate to turn off the water [WS+10].
- Show-me (Picture 8): Kappel and Grechenig developed a shower water meter [KG09] that displays the amount of water used during one shower in the form of LEDs assembled on a stick.
- Shower calendar (Picture 11): It measures water consumption at the armature and displays the consumption per resident on a display outside the shower [LH+11]. Further development is intended to visualize the consumption on the shower screen.
- Invetech system (Picture 13): A measuring device that captures flow-rate, shower volume and additional information for increasing the awareness of water usage in the shower. Besides the display, audible interaction is possible. Just like the shower timers, residents are being gradually motivated to spend less time showering.

All of the above-mentioned devices are battery-powered. A novel approach towards powering water sensors has been explored by DAIAD-partner Amphiro AG [SF11], a spin-off of ETH Zurich. The company has developed a battery-less water sensor, integrated with a micro-generator providing power from water motion, thus eliminating the need for an additional power source (Picture 12). The self-powered measurement device is packaged in the form of a cylinder about the size of a wine cork and can be easily installed by consumers in the bath faucet (plug and play). The device communicates with the display through a proprietary, very low power infrared protocol. A recently introduced generation of this product, amphiro a1 [Am13], integrates the sensor and the display directly to the shower tap, maintaining battery-less operation. It is integrated between standardized shower hoses and showerheads and measures various consumption data (Picture 14). Moreover, the devices do not require the user to perform a calibration process. Specific information on the consumption is visualized in real-time referring to the current shower [TT+13]. However, the solution at its current stage can only measure water and energy consumption, does not contain a communication module, and thus does not provide consumption data to third party systems for further analysis.

The major drawback for most of the sensing technologies comes from their power requirements or the need to calibrate them. In contrast to electricity meters, water-sensing equipment would require either a battery or a dedicated connection to the electrical grid; both is not feasible for mass adoption in the residential sector for fixture-level devices. The prototypes and product from Amphiro AG are energy-autarkic and do not require a calibration process, but thus far have no network capability.

1.3. Multiple fixture sensing (centralized systems)

Most centralized systems are based on advanced smart water meters that automatically and electronically capture, collect, and communicate water consumption almost in real-time. Time-scaled recorded flow information is then collected through a wireless or wired connection. The obvious benefits include the repurposing of an existing or anyhow emerging infrastructure and thus the reduced cost and complexity for offering add-on services. Beyond the arguments concerning smart water meter rollouts, a number of technical constraints have to be taken into account. While data collected from smart water meters is real-time at the point of measurement, the measurements are aggregated and then forwarded for processing, at best at 10 min intervals. Yet real-time feedback that influences an action while it is occurring even requires latencies *in the range of seconds*. Frequent transmissions on the order of 1/10 min are not practical even for expensive lithium-battery-powered meters due to the power consumption of the radio module, thus energy-harvesting techniques would be required. Furthermore, to extract additional knowledge concerning fixture consumption, flow trace analysis must be performed. Flow trace analysis so far is not a completely automated process, requiring manual data entry. In addition, although flow-trace analysis has been shown to be capable of classifying extractions at the fixture level at least in laboratory setting, it is still difficult to distinguish hot and cold water extractions and thus to provide energy-related information. Furthermore, centralized disaggregation systems require dedicated training periods in each household they are installed, which must be repeated whenever changes to one of the outlet are made – e.g., after a new shower head is installed, a new washing machine in purchased, or an aerator is built in a faucet. This considerably limits the applicability of the otherwise elegant approach to fine-grained consumption information from a central measurement device. Moreover, while consumption level data located at a central server might aid supplier processes, strong feedback interventions typically base on in-situ feedback at the point of use. Therefore, faucet-level displays would be required anyhow, and these devices might take over measurement capabilities as well.

Given the advantages and disadvantages of both centralized and decentralized systems, we included both types in this state-of-the-art analysis. The following section complements the discussion with additional aspects regarding centralized disaggregation approaches.

1.3.1. Vibration sensors

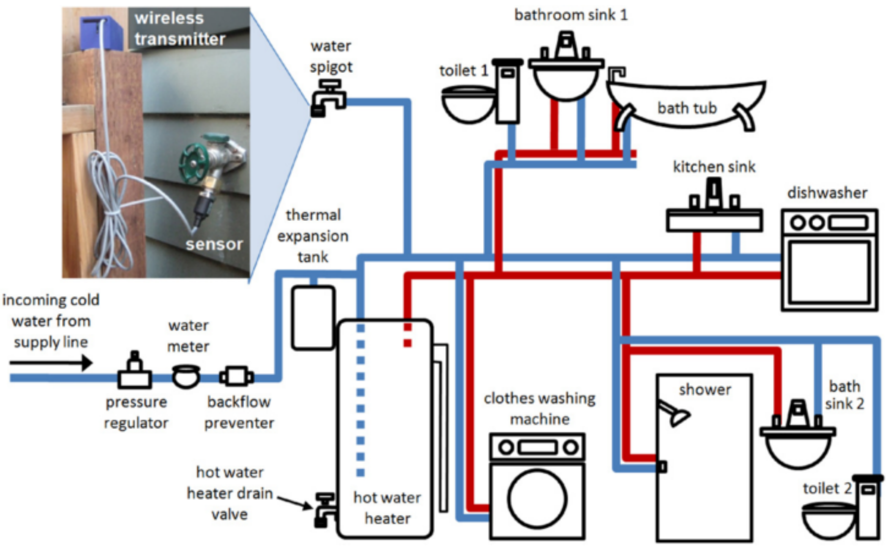
In contrast to the commonly used sensing mechanisms described in the section of conventional water metering other measuring techniques such as accelerometers have been explored and documented in the literature. Evans *et al*. showed in a laboratory environment that *accelerometers* mounted on the exterior of water pipes have a strong deterministic relationship to water flow rate [EB+04], but this is highly sensitive to pipe diameter, material, and configuration; the technology is thus not applicable in the field. The accelerometer senses pipe vibrations induced by the translation of pressure fluctuations in the fluid. Kim *et al*. proposed using a home's existing aggregate water flow meter together with a network of accelerometers on pipes to infer flow rates throughout a home [KS+08], achieving higher precision yet not overcoming the practical problems prohibiting application in the field.

1.3.2. Noise sensors

Fogarty *et al.* [FA+06] used microphones pressed against the exterior of a single home's major water pipes to demonstrate recognition based on patterns of water use (e.g., the series of fill cycles associated with a dishwasher). The hardware consists of a microprocessor, a microphone as sensor and supports wireless communication. As a prototype, the sensor board was directly attached with a tape to the pipe. Every two seconds the sensor captures 1,000 microphone readings and stores them in a memory. Aggregated data is then transferred to a laptop running the logging software. Trained algorithms detect water use patterns. Drawbacks of this work include the inability to differentiate among multiple instances of similar fixtures and concurrent activities, vulnerability to interference from ambient noise, and lack of water flow estimation.

1.3.3. Pressure sensors

Froehlich *et al.* [FL+09][LF+12] developed HydroSense, a low-cost pressure-based sensor that automatically determines water usage activity and flow from a single non-intrusive installation point, by continuously analyzing pressure in a residential water infrastructure (Picture 15). Identification of water fixtures is based on the unique pressure waves that propagate to the sensor when valves are opened or closed. The amount of water being used at a fixture is estimated from the magnitude of the resulting pressure drop within the water infrastructure. An experimental evaluation indicated that the proposed algorithms identified fixture events with 98% accuracy and estimated water usage with error rates comparable to empirical studies of traditional utility-supplied water meters. However, the proposed approach requires an external power source to operate. Moreover, the technique requires a time consuming and sophisticated learning phase to train the system in order to determine which pressure pattern belongs to which outlet. Changes in pressure drop are likely to lead to different pressure signatures, which dramatically reduces accuracy. Moreover, even smaller changes of the infrastructure (e.g. new shower head in the shower, new washing machine) would require a new training phase. While the approach itself is interesting from an academic perspective, it cannot be foreseen when and how the challenges of applying the results at a commercial scale can be overcome.



Picture 15: Structure of the HydroSense architecture

Disaggregated sensing reveals the potential to gain single fixture consumption data by the use of artificial intelligence such as expert systems. However, the machine learning driven methods still involve accuracy and training problems. Sensors for pressure, noise and vibration sensing need a specific minimal flow rate to well function and to deliver adequate data for the software analysis. The sensed data can easily be combined with beforehand surveyed behavioral information to increase precision.

1.4. Power supply

The reason why smart, energy-aware faucets and related data-driven conservation mechanisms are not in widespread use is simple to explain yet difficult to overcome: Implementing IT-based systems in the water domain is difficult, as existing solutions require an electrical power supply to operate. Battery-powered solutions are cumbersome since batteries need to be exchanged and thus lack acceptance especially among both "*green*" consumers due to environmental concerns and commercial services providers that are aware of high cost of replacing batteries in the field. Moreover, the inclusion of battery compartments contradicts the design imperative of the manufacturers who demand sleek surfaces. Connecting faucets to the electrical grid is only feasible in industrial settings and the highest-priced consumer segment (e.g., Grohe's Ondus [Gr13]) as it requires certified craftsmen to perform installations in wet environments. In the commercial sector (e.g., airports, restaurants), electronic faucets that use generators powered by water flow are already commercially available, yet the generators have the size of a large tea cup and thus cannot be integrated in bathroom and kitchen faucets for the mass market. In fact, attempts to downscale such systems have failed to produce products that are cheap and practical enough to make their way into actual products (see, for example, the project PowerFLUID [FS+09]). As a consequence, systems to provide accurate, real-time, and highly granular data of domestic water consumption for individual outlets (e.g., shower, kitchen) are currently hardly available, despite the interesting market environment for water flow sensing equipment.

1.4.1. Wired water meters

Centralized and distributed sensing technologies require electrical power to operate. Despite high installation cost, wired installations are attractive for reliability and security reasons for main inlets in households. There, wires are often used for both, power supply and data communication (e.g., for wired M-Bus communications). However, wired solutions inhibit complex installation problems whenever many extraction points need to be observed, and are not suited for agile environments with changing devices. Thus, they do not seem as the right fit for easy retrofit solutions as put forth by the DAIAD sensing vision.

1.4.2. Battery systems

Due to the complexity and cost of installing wired water meters, batteries and RF communication principles are gaining popularity. Batteries power most of the end-customer devices as discussed in section 1.2.2. These are mainly operated with lithium-ion batteries, such as the shower timer *waterpebble* for fixture sensing that uses a lithium battery with 3V or the WaiTEK shower monitor. Metering systems that communicate to a gateway mostly integrate ZigBee, Bluetooth, wireless M-Bus, or proprietary protocols. For example, HydroSense is battery powered to sense water pressure and transmit data via Bluetooth. Current problems of chemical energy storage represent the lifetime. For example, Fogarty *et al.* [CL+10] operate D-cell batteries for each noise sensor. The sensors run for six weeks over which sensor activity, memory, and wireless data transfer are enabled. The lifetime is rather short and illustrates the dependencies of such systems of the residents to change or recharge batteries during operation. Other prototypes succeed in providing a longer operating time: Lee *et al.* [LE+08] implemented a wireless digital water meter with a hall sensor for registering the water consumption combined with a ZigBee data transfer protocol. The focus lies on low power consumption for a long lifetime of the batteries (3V, 3,000 mAh) for about 8 years. Additionally, industry water meters such as SITRANS[2] from Siemens are supposed to have a 6 until 10 year lifetime (12-24V). The batteries, however, are often more expensive than the meters and the communication modules together and raise environmental concerns during production and disposal.

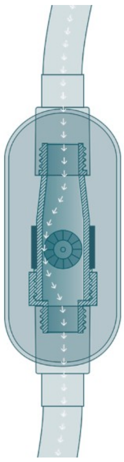
[2] http://www.automation.siemens.com

1.4.3. Self-powered water meters

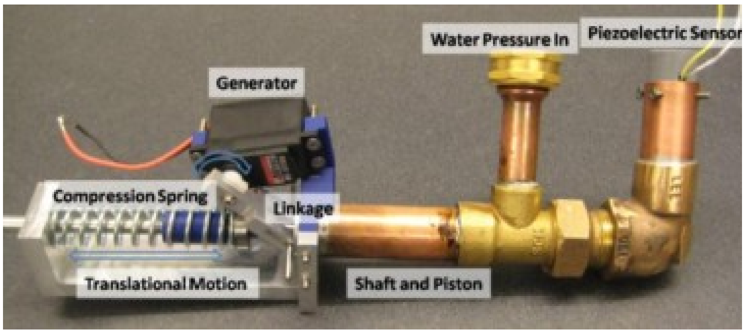
Self-powered water meters overcome the problems of complex installation in wet environments and energy storage dependencies. They can act autarkic and do not need any resident action such as battery replacement. Self-powered devices usually harvest energy from solar energy, ambient vibrations, or inductive coupling to RFID readers for data transfer. Water flow sensing opens new possibilities for harvesting energy and operating self-powered water meters.

Fröhlich *et al.* optimized their HydroSense water-sensing meter with a self-powered wireless version called WATTR for single fixtures [CL+10]. The sensor node collects and transmits water pressure transients in residential plumbing. During the sensing activity the device harvests energy from piezoelectric power for the operation of the device. It generates up to 15 mJ of energy and transmits data every 8 seconds. The prototype is shown in picture 17.

Amphiro a1 is self-powered through a generator module operating through the water flow from the shower hose to the showerhead. It uses mechanical metering technology from inferential-meters with an impeller. When water is passing through the device, the vane generates enough electricity for sensing the water flow and operating a display delivering real-time feedback. The device is sketched in picture 16.



Picture 16: Self-powering mechanism of amphiro a1



Picture 17: WATTR mechanical harvester and sensor

1.5. Connectivity

Network connectivity of water meters is of concern for both water utility companies and customers. Since meters are often installed directly inside the apartments, physical meter readings require access to the private space and thus a high degree of organization and coordination with the customers (which have to wait for the service person on a to-be-agreed-on date) for executing yearly or even more frequent readings. Utility staff or the customer herself needs to read the register and communicate the status to the billing department of the utility. Due to the labor intensitiy, manual meter readings are cost and time intense. Changes of tenants require additional, costly readings outside the regular billing intervals. Thus, considerable effort has been made by industry and researchers to establish remote meter reading processes.

Generally, Satterfield & Bhardwaj differentiate direct, remote, and automatic (or advanced) reading [GS+11]. Direct (or manual) reading refers to a service person going from door to door to read out meter registers. The employee enters the register status manually on a sheet of paper or on a laptop. In more advanced settings, the plausibility of the meter reading is checked in real-time via a networked terminal of the operator in order to reduce the likelihood of wrong data entry and subsequent expensive settlement processes.

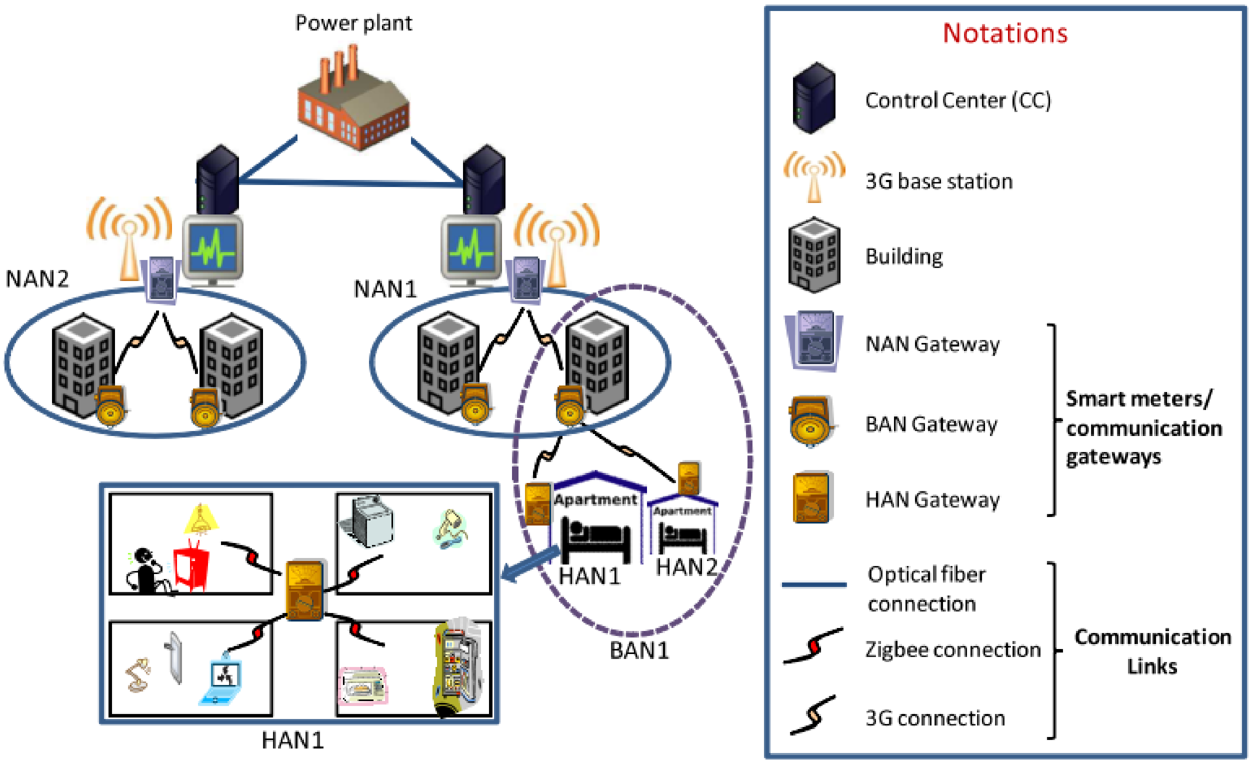
Remote reading describes the automatic transfer of the register status from the meter to a separate output station (terminal). For this approach, meter and terminal are typically connected via infrared. Automatic reading represents an enhanced version of remote reading in which no person is required to initiate the reading process locally. Data transfer is typically done via radio transmission or power line communication (often using a smart electricity meter). This setting has lower variable costs (cost per reading), and the service personnel does not require physical access to the apartment. Due to better transmission technologies, the transmission range of RF meters increased dramatically over the last years, and the operator just drives by the buildings to gather the required data. Thus, meters can be read out more often, and more granular data becomes available to customers and utilities.

Advanced meter reading (AMR) represents the next step of consumption data retrieval [De06]. AMR is also referred to as smart metering. AMR offers automatic, high-resolution data acquisition by a central information system and subsequent analysis. Based on AMR data, dynamic pricing and more advanced water conservation campaigns become possible. For AMR, various communication technologies are used; an overview of the common approaches and their technical specifications is presented in Table 1 [GS+11][PK+10].

Technology	Spectrum	Data Rate	Coverage Range	Applications	Limitations
GSM	900 - 1800 MHz	Up to 14.4 Kbps	1-10 km	AMI, Demand Response, HAN	Low data rates
GPRS	900 - 1800 MHz	Up to 170 kbps	1-10 km	AMI, Demand Response, HAN	Low data rates
WLAN	2,4 GHz	1-54 Mbps	100 m	HAN, PAN	Reliability, availability
WiMAX	2,5-66 GHz	Up to 70Mbps	50 km	WAN	WiMAX towers are expensive
3G /4G	1.92-1.98 / GHz 2.11-2.17 GHz (licensed)	384 Kbps-2 Mbps	1-10 km	AMI, Demand Response, HAN	Costly spectrum fees
PLC	1-30 MHz	2-3 Mbps	1-3 km	AMI, Fraud Detection	Harsh, noisy channel environment
ZigBee long range	868 - 915 MHz, 2,4 Ghz	250 Kbps	10-100 m (more in long range versions)	AMI, HAN	Low data rates, short range
Bluetooth	2,4-2,4835 GHz	721 Kbps	1-100m	HAN	Weak security

Table 1: Overview of communication technologies for water sensing

For the DAIAD project, it is relevant to analyze both short and long range connectivity technologies for the final integration of all devices. For the realization of a system that makes “water sensing available to everyone”, especially low-power and low-cost technologies matter, and adequate mechanisms for centralized and distributed water sensing need to be adapted.



Picture 18: Smart Grid communication architecture [FF+11]

Generally, various players and network technologies can be identified as vital for interconnected, intelligent devices and meters in a smart grid and metering environment [FF+11]. For water metering, the differentiation of Neighborhood Area Network (NAN), Building Area Network (BAN), and Home Area Network (HAN) is applicable (Picture 18). HANs can integrate communication for fixture sensing and aggregate all data for a residence. Building and Neighborhood Area Networks centralize and aggregate fixture and household level consumption data and transfer it to the utility. Centralized water sensing is relevant in the context of BANs and NANs. For these networks, there exist no common standards that are widely established. Therefore, various technologies are taken into consideration and their potential to aggregate single fixture data and to transfer household or building wise information to the utilities is discussed in the following.

1.5.1. Power Line Communication (PLC)

Power line communication (PLC) refers to – as the name suggests – data transfer over the electrical grid [GS+11][PK+10][CL+10]. The technology is widely used for baby phones and in-house networks. Utilities operate devices and meters with day and night tariffs with the help of (a rather old version of) PLC. Generally, the already existing infrastructure entails low costs, yet the bandwidth is rather small. The field of application is therefore limited to networks within a building with speeds of typically around 28 Mbps and low speed, high latency connections over larger distances with typically 20 kbps per line. For meter connection scenarios, meters are connected to a data concentrator that aggregates the consumption data from typically 50 to 100 devices and transfers the data via GPRS/GSM, DSL, or optical fiber to the utility's data center. The low bandwidth is often acceptable for settings with monthly or daily meter readings, and utilities often favor PLC over direct GSM-meter connections since the PLC-network is under their control and no costs from external data communication providers occur.

1.5.2. Digital Subscriber Line (DSL)

Digital Subscriber Lines (DSL) represent transmission technology using telephone lines with frequencies greater than 1 MHz. Again, the already existing infrastructure represents a vital benefit for the usage of this communication technology for (smart) water metering. The usage is restricted by the potential of down time and low reliability of DSL technology. Additionally, distance dependencies provide that the technology is only utilizable by certain customers in high-density areas.

1.5.3. Global System for Mobile/General Packet Radio Service (GSM/GPRS)

The Global System for Mobile (GSM) serves as an attractive communication technology due to a widespread, existing infrastructure. Many telecommunication providers are eager to support smart metering in their infrastructure, and vendors integrate General Packet Radio Service (GPRS) modules in there products (see, for example, Itron's SENTINEL [GS+11]). The technology is mainly used for current centralized sensing due to the possibility of transferring larger data volumes. Meters aggregate and communicate the data to servers and systems of the utilities (see section 1.6.2). Security issues for data transmission can be addressed quickly because of the existing, elaborated infrastructure. Moreover, cellular networks can also support distributed communication in in-home networks. The 3rd and 4th generation of cellular technology offers continuous improvement for higher data rates and better quality of service [PK+10]. However, relatively high costs of communication and connectivity issues in the basement of houses – where meters are often installed – still constitute challenges that need to be resolved.

1.5.4. Worldwide Interoperability for Microwave Access (WiMAX)

The Worldwide Interoperability for Microwave Access (WiMAX) technology is standardized since 2001 for wide-band communication structures of 10-66 GHz [PK+10]. The data rate is currently up to 70Mbps with a distance covered of almost 50km (with an increasing distance, the network speed decreases). Future applications include AMR. With its performance, WiMAX technology may automate the meter reading for entire cities using very few receiving towers. The radio frequency hardware for WiMAX tower still is expensive, yet the outlook is very promising.

1.5.5. Wireless Local Area Network (WLAN)

Wireless Local Area Networks (WLAN) enable high data rate applications with a very extensive geographic coverage [FF+11]. Currently, the technology is mainly used for in-home communication. Many residences are already equipped and form a developed infrastructure. However, devices using WLAN are still characterized by high power consumption, though low power implementations are under way. Yet decentralized metering solutions might not be cost-effective for battery-powered systems. Moreover, most utilities would not rely on their customers' infrastructure (i.e., their WLANs) for integrating them into critical business processes such as billing. For non-critical point of use devices, WLAN may be the technology of choice.

1.5.6. Bluetooth

Bluetooth supports wireless Personal Area Networks (PAN). The technology is well suited for low power and short-range communication. Currently, it is often used for voice, data, and audio transmission [FF+11]. The technology enables point-to-point or point-to-multipoint communication over 100 m (e.g., the WATTR [CL+10] uses Bluetooth to send data from single fixtures to the HydroSense system). Currently, security issues represent a disadvantage in comparison to other standards. The current version, Bluetooth 4.0/4.1, concentrates on low energy consumption and optimizes connection time and distance. Moreover, sleeping modes and synchronized transmission is included.

1.5.7. ZigBee

Another widespread wireless communication technology is ZigBee. It has been designed with a special focus on controlling lighting, energy monitoring, home automation, and AMR. Based on the IEEE 802.15.4-2006 standard, ZigBee encompasses a few protocols for high-level communication supported by small, low-power digital radio devices. Typical implementations offer transmission rates of 20-250 kbps and transmission ranges of up to 400 m [GG09]. For that reason, sensors and transceivers can be used in small residencies up to larger customers such as companies or public services. Up to 65,000 devices can be connected to one system that allows for a rather broad employment of the technology. To date, ZigBee is frequently deployed for Wireless Sensor Networks (WSN), and lots of experience exists when integrating the protocols. Another positive aspect is that meter vendors such as Landis+Gyr, Itron, and Elster announced to use ZigBee in their products [GS+11].

ZigBee modules are very robust, simple, and come at low cost [PK+10]. With these characteristics, ZigBee is also suited for decentralized metering at the fixture level. However, interferences with other protocols that share the same frequency band (mostly WLAN) constitute a disadvantage. Moreover, stated transmissions ranges apply to applications in US homes and are not met when the technology is deployed in brick-and-mortar homes in Europe. Finally, corruption of the communication channel constitutes another considerable challenge regarding the protocol in the security context.

1.5.8. Meter-Bus (M-Bus)

The Meter (M)-Bus represents a European standard for remote data reading for utility meters and home automation devices [GS+11][CD10]. A master device connects various meters, sensors, and actuators (slaves). It forms a so-called segment and collects consumption data periodically. A wired and a wireless version exist. For the wired version, the master can provide power to the slaves, what considerably reduces the complexity during installation and operation. In every system or segment of master and slaves exists a limitation regarding the number of devices (around 250). For bigger systems, repeaters are required. A few kilometers of connection distance can be covered. Just like ZigBee, the focus lies on simple, low-cost, and low powered devices. The system can be used for the aggregation of decentralized smart water metering to gather consumption data from various fixtures.

1.6. Smart home

Ubiquitous computing and the “Internet of Things” describe a world where “everyday objects that have not previously seemed electronic at all are starting to be online with embedded sensors and microprocessors, communicating with each other and the Internet” [SW12]. Such systems can, for example, be used in buildings, clothing, household appliances, and other commodities.

Future domestic application fields of ubiquitous computing are closely related to the smart home domain. The vision of smart homes describes homes with ubiquitous computing applications that are connected with each other. Examples include heating and ventilation systems that work together to ensure a comfortable room climate. Other smart household appliances including refrigerators, washing machines, and dishwashers, that may organize themselves such that they preferably operate when electricity prices are low. More contemporary applications include presence detection that controls lighting and media devices depending on the persons in close proximity. Presence detection may also serve as a security system that informs residents via SMS about entries and exits to their apartment. In order to integrate all these data receiving and data providing entities, communication standards such as ZigBee and WLAN are needed to support interaction and communication. Another important aspect for smart homes includes sensors for retrieving environmental data (such as temperature, humidity, noise, pressure, etc.), data on objects (position, movement), resource consumption (electricity, water, gas, etc.), and last but not least actuators (e.g., for opening windows, turning appliances on and off, etc.). In this context, smart meters can provide data on consumption, user behavior, and appliance status. The role of smart metering for home automation is outlined in the following section.

1.6.1. Smart appliances and home automation systems

In smart homes, a large number of appliances are equipped with sensors, actuators, and communication modules; they become part of an infrastructure where physical objects interact with each other. In this context, smart water meters that "automatically and electronically capture, collect and communicate up-to-date water usage readings on a real-time […] basis" [SW+10] are yet another class of smart objects within the home infrastructure.

This meter data can be used in various ways. First of all, it can help to promote conservation behavior among the consumers. Second, demand side management strategies that built on the data can help to automate appliances depending on tariffs that in turn may depend on demand peaks. For both, demand curtailment and demand shifting, in-home displays and other means of visualization (e.g., via smart phones) play an important role. Third, data from smart water metering can help to go beyond water and energy conservation. Data traces may help to improve presence information and reveal user habits (e.g., on time of activity, number of visitors, etc.). Thus, the data traces can be both, helpful to built better automation services and harmful when the data is not used in the will of the consumers the data is related to. Yet the benefit of smart water meters comes from integrating various systems with each other within a smart home, so open standards should be promoted.

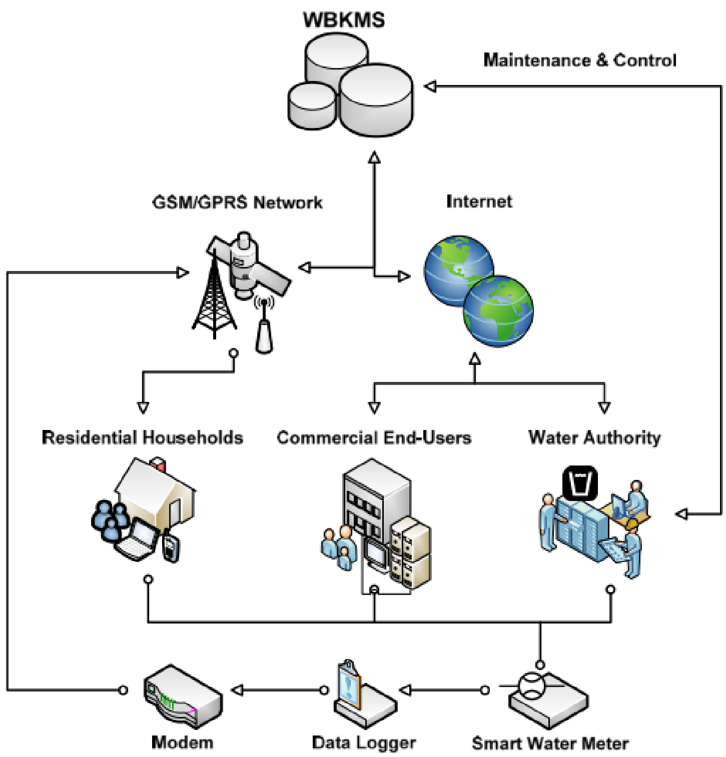
1.6.2. Smart home integration systems

The section describes three exemplary smart home integration systems for smart water meters on various levels. In the first setting, the smart water meter enhances knowledge of all participants only on water consumption. In the second case, the integration perspective includes other resources such as gas and electricity. And finally, the smart water meter is a part of a whole citywide system incorporating additionally information in order to promote sustainability and comfort for the citizens.

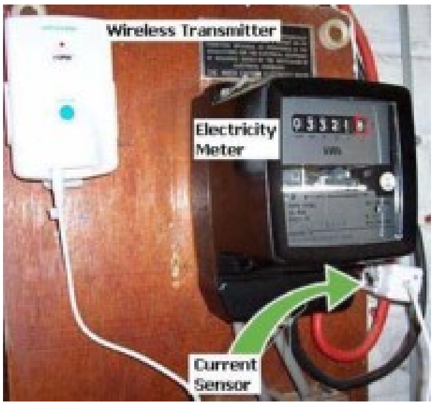
1.6.2.1. Web-based knowledge management systems

The major goal of the web-based knowledge management system (WBKMS) discussed in this subsection is to promote water savings by means of an integrated smart water meter infrastructure [SW+09][SW+10] (Picture 20). An emphasis is put on data analysis for various users. The system incorporates real-time water consumption sensing. Consumption data is collected on the level of individual households (not individual faucets). A knowledge repository stores consumption data for further analysis. Customers, water utilities, and government organizations may access the data online (left aside measures to protect consumer privacy). Customers can retrieve their own daily, weekly, and monthly water consumption. Additionally to historical data, a pattern analysis and an activity specific overview is included. Moreover, in the billing section, the system delivers hourly cost information. For certain events, such as excessive water use, online alarms may be set. Utilities are enabled to detect leaks faster and at higher levels of confidence. The optimization regarding leak detection leads to a more specific reaction and shorter intervention time. Early identification and repairing entails less water losses. Besides, the system provides household classification through pattern recognition. The combination with demographic data enhances the understanding of different water usage. Thus, provision planning and education campaigns can be tailored as well as optimized.

On the architectural level, the system uses well-known technologies such as GSM/GPRS for data acquisition. Smart water meters are connected to a data logger aggregating consumption data and delivering it, when needed, via GSM/GPRS to the WBKMS system. The stored data is processed for automated reports. Reports are delivered via web interfaces to the specific audience. Water authorities host and control the WBKMS system. Additional wireless sensors and matching algorithms for flow traces help to improve data quality and precision.



Picture 20: Structure of the WBKMS



Picture 19: Meter installation for Smart metering project

Picture 21: EcoMonitor

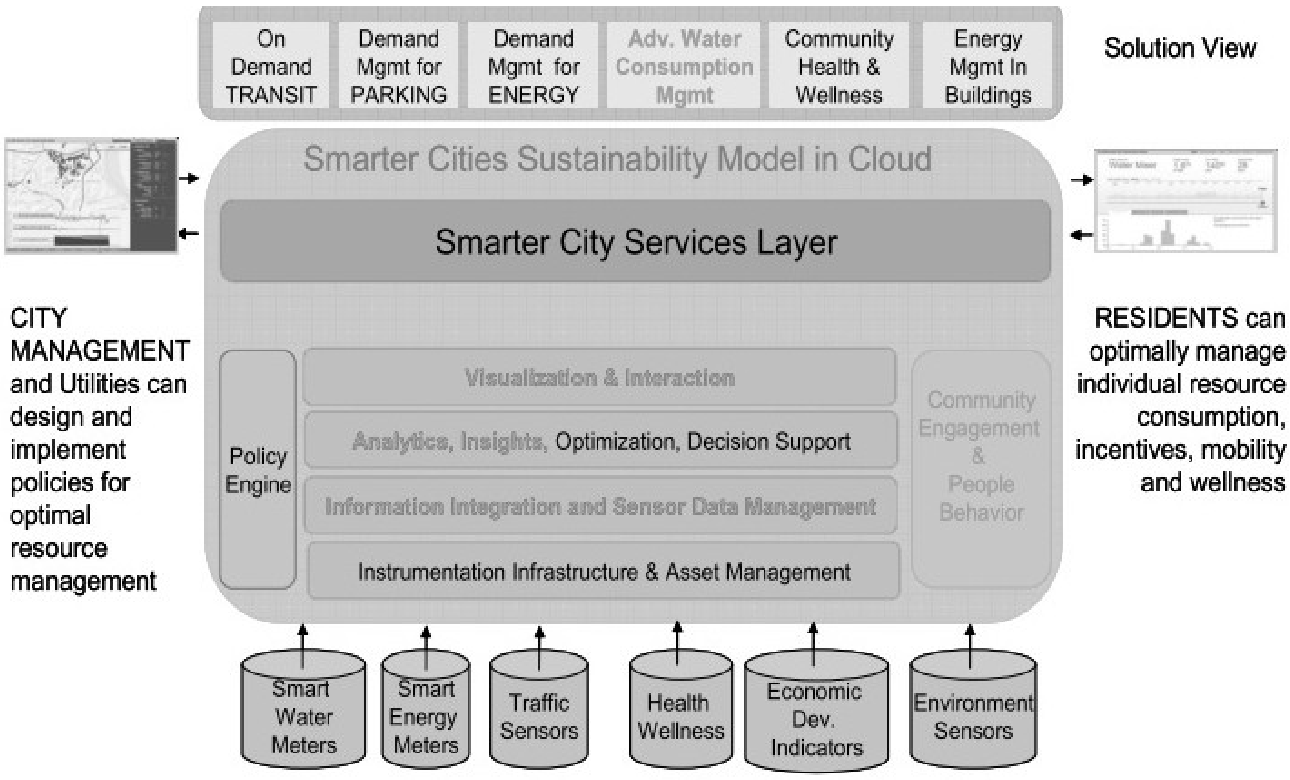
1.6.2.2. Integrating smart metering for all resources

The second project under review, the EcoMeter (Picture 21), integrates water with gas and electricity metering and displays the consumption on an in-home display. The metering system (Picture 19) is connected via an RF interface to a network management system. Daily consumption data is transferred via GPRS for further analysis. The EcoMeter uses PLC for data transfer to the in-home Display. It provides real-time as well as historical data on gas, water, and electricity consumption. It is also possible to access data via the Internet over a web browser. Moreover, service providers can set alarm levels and annotate texts to improve consumption feedback and to interact with the residents. However, there is only limited information available on the technical details of the system and the impact the system has on user behavior.

1.6.2.3. IBM smarter city sustainability model for the Dubuque water pilot

The Smarter City Sustainability Model (SCSM) represents a holistic structure for integrating smart water metering with other domains concerning the citizen's quality of life. The domains embrace energy consumption, traffic, health, economy, and the environment [NL+11]. The system uses the concept of cloud computing and aggregates data from various sensors: smart water and energy meters, traffic sensors, economic deviance indicators, and many more environmental sensors. Further data such as weather or demographic data can be incorporated. City management and residents have access to the information. For each user group, specific services are offered. The analysis and deployment of data warehouse and data mining activities deliver citywide metrics, reports (such as for anomaly detection, forecasting or trending), alerts, and consumption feedback. City planners, utilities, and policy makers use the metrics and reports for improving public infrastructure. Residents are enabled to optimize their resource consumption. The SCSM includes a special layer for visualization and interaction. Accordingly to Naphade *et al.* [NL+11] an incentive design for customers including social networking functions and customizable interfaces, enhances behavioral change of customers. The sense of community is an important aspect of the SCSM. Novel aspects of the SCSM in comparison to the WBKMS include the integration of various data sets and on the system and interface development for customers. Nephade *et al.* [NL+11] take into consideration that energy conservation needs to be motivated by the system.

The next chapter concentrates on interventions and user interfaces used in conservation campaigns.



Picture 22: The structure of the SCSM (source: IBM)

2. Interventions and Interfaces for water consumption

Researchers have conducted a plethora of studies to investigate the effects of behavioral interventions on resource consumption [AS+05]. They have aimed at both changing the context in which consumption decisions are made (e.g., offering rewards that render pro-environmental choices more attractive [SV09][Os97]) and at targeting individuals' perceptions, preferences, and abilities to induce eco-friendly behavior [Al82][PS+03][St08]. In this context, interventions that refer to a specific situation, state of knowledge, or feeling, appear to yield higher savings than less specific interventions [PK+11]. Abrahamse *et al.* point out that tailored information, such as home audits conducted by energy saving experts, is more effective than general advice [AS+07]. However, to generate such tailored interventions is time consuming and therefore does not scale when targeting larger numbers of households, as one must gather and process the underlying data, and translate the results in a form that is both easy to understand and motivating for the specific addressee. Traditional powerful intervention strategies (e.g., audits, one-to-one consulting, etc.) suffer from limitations similar to those faced by brick-and-mortar retail, including the classical trade-off between reach (i.e., "*the ability to connect with a large number of actors*" [EW99]), and richness (i.e., "*the ability of information to change understanding within a time interval*" [DL84]). In this context, authors have stressed the potential of ICT systems to resolve this conflict and to establish cost-efficient energy efficiency services with the aid of automated data processing, personalization, and immediate feedback [OH09]. Various authors have presented reviews of the early works in this new but quickly evolving research stream (e.g., [Mo09][IM+10][BA11][Me10][JW+11]). A number of related contributions exists in adjacent domains, including human-computer interaction, energy policy, and social psychology research.

In order to study interventions that curb water and energy consumption, this section provides a detailed overview on the psychological concepts governing behavior, motivation, and decision-making that are applicable to the system envisioned in DAIAD. Additionally, the section presents empirical evidence on the effects of ICT-based interventions on water and energy demand. An overview of feedback devices for water consumption follows, and lessons learned from related approaches in the electricity domain are presented to guide the design of interventions and interfaces within DAIAD.

2.1. Psychological concepts

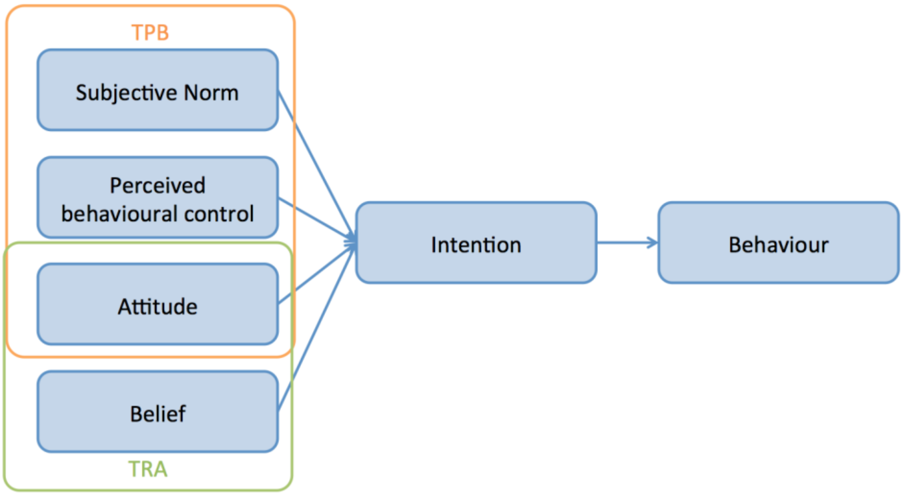
Energy conservation campaigns in the domestic sector frequently aim at changing *day-to-day consumption behavior* of households, be it to curb overall demand or to reduce demand peaks. To achieve this, early interventions traditionally focused on informing individuals about their personal resource usage in a rather technical way (mostly targeting electricity), and often described general ways to reach a target behavior. Additionally, many of such measures pointed out potential monetary savings from changing demand. However, these campaigns mostly led to a rather limited engagement among the addressees, and therefore contemporary interventions increasingly apply more sophisticated behavioral *nudges* [AM10] to motivate a target behavior. These nudges tend to implicate functions to which individuals are more susceptible. A large number of theories and models exist to explain the underlying behavioral mechanisms of these nudges. These theories help to improve and select suitable approaches and consequently aid intervention and interface design. We therefore outline important theories below.

2.1.1. Rational-choice model and bounded rationality

Most basic theories characterize human behavior as a striving for maximizing personal benefits. So, accordingly to the rational-choice model, behavior is guided by costs, benefits, and individual preferences [LS+13]. The moment a human being is confronted with a decision between two alternatives, she is in a decision-making situation. Depending on the available information and individual preferences, the alternative resulting in the highest expected utility will be chosen. The decision-making process is based on cognitive deliberations [Ha07][SW+13]. These deliberations are based on all information available to the individual. It is supposed that the individual disposes over (a) all relevant information required for the decision, and (b) the necessary cognitive capacity to deliberate on all alternatives. In 1955, Simon [Si55] already expressed serious doubts about the applicability of this behavioral model when applied to the business domain. He describes the rational-choice model as an *approximation* of actual behavior. To begin, an individual's perception is limited by the cognitive capacity to detect all eligible alternatives. On this account, just a subset of all alternatives is considered. Secondly, the available information for the decision-making process is often restricted. The individual decides on a behavior without possessing all relevant information such as the actual outcome or the probability of occurrence. Lastly, individual preferences may change over time. So, the perceived utility at the moment of the decision might change as the preference system evolves. All in all, with limited information and confined cognitive capacity, individuals tend to apply "*mental shortcuts such as habits, routines, cues and heuristics*" [SW+13]. This partly explains why traditional interventions that solely built on rational choice of the actors often leads to unsatisfying energy saving effects. To achieve larger effects, social psychological insights are relevant to really understand human behavior.

2.1.2. Intentions

The theory of reasoned action (TRA) and the theory of planned behavior (TPB) entitle two frequently used models for "understanding, predicting, and changing human social behavior" [Aj12]. The general concept is about behavioral intentions. They "*constitute a willful state of choice where one makes a self-implicated statement as to a future course of action*" [LM+83]. Thus, intentions predict actual decisions for a future behavior or actions performed by the individual herself [Br01]. Various antecedents precede intentions depending on the prediction or explanation model (see Picture 23). The TRA includes *beliefs* and *attitudes* as antecedents for intention formation. Attitude is seen as a "strong predictor" for intentions [VV+13] in the way that it represents "*a learned predisposition to respond in a consistently favorable or unfavorable manner with respect to a given object*" or behavior [FA75]. So, an attitude is based on feelings and evaluations, which might be formed subconsciously. Beliefs represent an individual's opinions and knowledge depending on the information available about the behavior. Ajzen [Aj91] enhances the model to the corresponding state-of-art (TPB) by including the *perceived control* that describes a person's perception towards her ability to perform the action. Moreover, social pressure - also called subjective norm - influences the intention for a specific behavior.



Picture 23: The TRA and the TPB

Depending on the actual theory (TPB or TRA), behavioral change is explained by modifying the influence of the beforehand mentioned antecedents. Attitudes and beliefs towards an object, person, or situation have to be altered for guiding a certain intention and resulting behavior. The TPB enlarges the perspective and includes directly social pressure to the influential constructs of behavior.

Altogether, the theories also represent a simplified perspective on human behavior [Lu93]. Other factors, also called situational factors (such as price, knowledge, and income), are marginally considered. The context of any decision is only treated as exogenous. Some studies prove high predictive value whereas others show poor correlation between the TRA and TPB constructs. For example, individuals confronted with new technologies might be unequally influenced by perceived behavioral control depending on their age. In that context, Triandis [Tr77] develops a model reinforcing the influence of emotional aspects and habit. The belief construct recedes in favor of social and affective constructs. Behavioral habits (heuristics) are also included at the same level as intentions for explaining a certain behavior. For that reason, the model incorporates unaccounted constructs of the rational-choice theory and the TPB/TRA. More information on TRA and TPB, as applied in the context of Water Demand Management (WDM), is provided in section 1.

2.1.3. Theory of cognitive dissonance

The theory of cognitive dissonance describes an individual's striving for consistency between her actions and attitudes [Fe57][Ha07][Ra12]: A person experiences psychological discomfort in a situation she holds two or more conflicting beliefs or thoughts. The person is going to reduce the dissonance (also called frustration or disequilibrium) by changing the cognition or the behavior. For example, an individual who smokes learns about the bad effect of smoking. At that moment, two conflicting cognitions exist: On the one hand, she enjoys smoking, and on the other hand she thinks that smoking is bad for her health. In order to reduce the dissonance, the individual has two options. First of all, she can change her behavior in favor of the newly acquired knowledge. Or, she can modify the cognition by reinforcing the positive effects of smoking (e.g., higher tolerance to stressful situation).

Cognitive dissonance can also be applied as a persuasion method. It is supposed to induce a more permanent attitude change than external persuasion theories [DT+92], possibly due to addressing the self-concept of a person. The self-concept, or the way someone thinks about herself, induces attitudinal and behavioral changes. Intentions may change more frequently, yet the self-concept is more stable on the long term. So, the desire of individuals to reduce dissonance can be leveraged to trigger behavioral change of individuals.

Another insight from the adaption of cognitive dissonance theory in pro-environmental behavior research is the *spillover* effect [Th90][Ja05]. A pro-environmental attitude serves as a strong predictor for other behavior. For example, a person recycling trash tends to participate in energy or water conservation as well. Thus, the positive attitude to the environment is translated into other areas of life. However, there is also strong evidence for moral licensing; this counter-effect suggests that individuals feel entitled to act unethically after having done a good deed [TS13]. The ongoing discussion on which effect is the dominant one highlights the need for further research in this field.

2.1.4. Intrinsic and extrinsic motivation

In pro-environmental research, intrinsic and extrinsic motivation is often distinguished for guiding individual behavior in various situations. Both forms of motivation are also relevant in other domains including human resource management (employee motivation) and education. Extrinsic motivation describes the behavioral influence by external factors [BT06]. These factors encompass *reward or prohibition*-based motivation. In contrast, intrinsic motivation is self-induced by the individual.

Experimental settings have proven that intrinsic motivation leads to stronger behavioral reaction or interest, as many experiments have shown. For example, college students were asked to solve a puzzle. They were divided into two groups. The first one got paid a small amount of money for playing puzzle and the second one did not receive money for the task. It was observed that college students without financial compensation also played the puzzle in their "free-time" and showed greater interest in the task than those who have been rewarded monetarily. Other experiments revealed that non-rewarded individuals demonstrated better compliance on the long run. Even though, externally motivated individuals were more engaged at the beginning. Bénabour & Tirole [BT06] actually demonstrate that monetary rewards may have negative effects on behavior. Transferring a task to someone without promising monetary rewards also expresses trust in someone's abilities to finish the task regardless of extrinsic cues. Trust may lead to a greater outcome. So, intrinsic motivation is also about self-esteem and confidence. Overall, the influence of rewards very much depends on the nature of the task, the value system of an individual receiving the reward, and the relation she has towards the person who assigns the task. Therefore, both extrinsic and intrinsic motivations are relevant for influencing behavior, and case specific considerations are necessary.

2.1.5. Goal setting and defaults

Goals allude to reference points that encourage behavioral change for a future desired situation [LS+13]. These reference points can be self-set or imposed by others [LS+80]. The actual goal acceptance is crucial for the accomplishment. Acceptance is self-induced or influenced by the environment. Goal setting describes the conscious decision for an outcome. This guides an individual's actions persistently to make efforts to reach the desired outcome [SW+09]. Fishbach & Finkelstein [FF12] state that goal pursuing is much more persistent when receiving feedback. With the help of feedback, the progress of an individual is evaluated. The individual is enabled to adjust her efforts and focus on the right actions.

Goal choice (and consequently the outcome) can be influenced by using default goals, where a default is the option a user gets if she does not actively choose another one [LS+13]. The concept is often used in online marketing to set an anchor point to "pull" a potential customer in a direction that is favorable for the seller. In the energy conservation context, goal setting can be used to avoid non-ambitious goals, which would not trigger a strong engagement.

2.1.6. Social norms and normative feedback

The social science literature unambiguously highlights the power of social norms as an important factor that directs human behavior, and various reasons are given to explain the effect of social norms [Al11]. Social norms can be seen as rules and behavior patterns that are implicitly known and respected by the majority of a society's members. They may differ from one group to another and thereby express an affiliation to a group (e.g., eating with fork and knife or chop sticks). Socialization plays an important role for following or internalizing the norms of a group or a society. On the one hand, a particular behavior might be penalized or rewarded by the society. On the other hand, following norms and rules of a certain group enables the individual to maintain a specific image and signalizes a particular group affiliation. For example, an individual shows conformity to the norms of upper nobility with the goal to provide evidence that she belongs to that societal subgroup. However, conformity with particular norms might be also explained by a lack of knowledge. According to the norm activation model, pro-social behavior is influenced by norms [BS03][Sc77]. Referring to Schwartz, norms are the only direct influence factors on pro-social or altruistic behavior. So, pro-social and pro-environmental behavior depends on the norms and the moral obligation of individuals. Awareness of negative consequences increases the sense of responsibility.

Social norms are also supposed to gradually increase the influence of a subgroup's behavior to a wider public [MS99]. The effect is explained by an individual's tendency to participate in the activities of others. Thereby, the communication of social norms describing positive or desired behavior of a subgroup has the potential to influence actions of a larger group. This mechanism is also called normative feedback. Normative feedback is based on communicating norms to socially influence behavior [LS+11]. Descriptive and injunctive feedback can be distinguished. The first one informs individuals about the behavior of others ("*what most people do*"). Injunctive feedback includes a specific assessment of the receiver's performance (*approval or disapproval, "what other people think about the behavior of the feedback receiver"*).

2.1.7. Social feedback and social comparison

After the discussion on the implicit social influence on behavior, the next concept treats explicit impacts by other individuals. The explicit impact consists of direct social feedback of peers or society members. The underlying concept is called social facilitation. It describes the fact that a person performs better when peers or colleagues have the possibility to keep track and watch the individual during an action [SW+09]. The same effect is also induced vice versa, when the individual keeps track of her peers' behaviors. At the moment an individual recognizes how well others perform, she is motivated to perform better in order to achieve the same outcome. In order to do so, people tend to change their behavior accordingly to the behavior of their peers. This effect is called social comparison, peer-effect, or bandwagon effect.

As Yates & Aronson [YA83] outline, water saving suggestions for showering induces *much less conservation efforts* than an exemplary individual applying the suggestion in front of others.

2.1.8. Information framing

Individuals tend to judge an outcome of an action based on the way the related information is framed. A campaign that promises savings of 365 EUR per year typically triggers a much stronger engagement than promising savings of 1

EUR per day, for example, which clearly violates the rationality assumption. As a response to such "cognitive biases" of individuals, Yates & Aronson [YA83] recommend specific measures for visualizing water usage during informational campaigns. The explicit focus on the way a piece of information is put forth is called framing. It describes the action of reformulating information (the representation of the context) without changing its content [MM00]. Yates & Aronson suggest advertising losses rather than gains, for example when residents should be coerced to use a feedback device. Moreover, depending on the reference point, individuals tend to regard a financial loss as being more relevant to them than an equal financial gain [TK91]. This finding is in line with the suggestions of the prospect theory that describes, among other effects, the concept of loss aversion. Secondly, residents consider a *single price information* (*integrating all benefits*) much more attractive than complex information about energy consumption.

Another way of framing consists of finding alternative representations for the information that is to be communicated. As an example, pro-environmental information campaigns may visualize greenhouse gas emissions and water consumption with metaphorical representations: Instead of using complex units, black balloons may visualize the amount of greenhouse gasses and buckets depict the liters of water [St11].

2.1.9. Hedonic motivation and Gamification

Hedonic motivation is about performing a behavior because of a purely intrinsic driver: pleasure. This is the basis of the concept of gamification, who's formal description dates back to the end of the twentieth century. Other terms such as funware, playful design, or behavioral games are used interchangeable. Gamification unifies the trend of an increasing ubiquity of video games and the evolution of non-game products into entertainment and enjoyable ones. For that reason, interactive systems (including systems related to nutrition, health, education, productivity, finance, sustainability, and energy conservation [SW+09]) contained very early *serious games*. Their main goals are motivation and increased user retention by adding game-like aspects to non-game products. Adding a gamification function to a service or product embraces a system with rewards and reputations such as points, badges, and levels that cannot be traded in money or tangible products, in most cases. However, social norms, comparisons, and especially the reputation among peers are important drivers. Particularly, energy conservation programs may convey the message that "saving is fun" [GJ+11]. Gamberini *et al.*, for example, describe an eco-game for residents using smart-phones (ubiquitous gaming). In this game, users increase their awareness on energy consumption by passing through various levels that involve quizzes and explorative actions.

The following section outlines the effects of some of the aforementioned concepts. It also highlights the necessity to pay attention to very detailed aspects regarding their implementation in order to leverage their full potential to trigger behavioral change.

2.2. Empirical evidence from technology-based interventions and interfaces

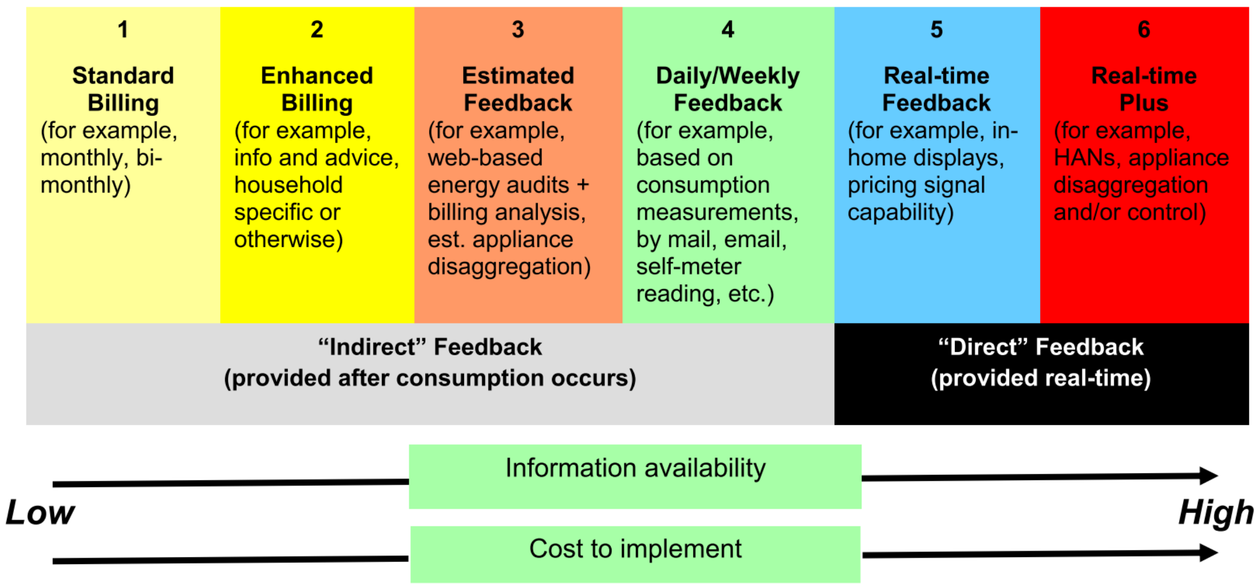
Feedback – or the provision of information on one's performance or state, often in relation to a target behavior – plays an important role for behavioral campaigns. In the context of their extensive literature review on residential electricity conservation, Neenan & Robinson [NR+09] identified various forms of feedback in the literature. They are summarized in Picture 24. One of the major differentiating factors is when the feedback is provided: *Direct feedback* follows and action immediately or is even provided the very moment an action is performed, whereas *indirect feedback* subsumes delayed information that – in most cases – is more difficult to attribute to a specific behavior.

Sub-categories of indirect feedback are:

- standard billing that provides yearly or bi-annual consumption data (1),
- enhanced billing that includes advices and supplementary information on conservation (2),
- estimated feedback that provides supplementary information, e.g., based on data entry in web portals for more detailed reporting and energy audits (3), and
- more frequent feedback, e.g., on a daily basis by email or self-meter reading (4).

Direct feedback is characterized by real-time information and virtually always utilized ICT (e.g., in-home displays,). Neenan & Robinson [NR+09] describe two subgroups:

- Real-time feedback, which provides aggregated information (e.g. current consumption as a sum of lighting and white goods), and
- Real-time feedback plus, which offers information on individual devices (e.g., for lighting and white goods separately).



Picture 24: Feedback Intervention types [NR09]

The next section outlines quantitative and qualitative evidence on the saving effects triggered by ICT-based interventions (estimated, daily/weekly, real-time, and real-time plus feedback). Based on a literature review, we also present interface designs and provide the connection to the psychological concepts identified above.

2.2.1. Quantitative evidence

Early ICT-related studies have primarily focused on the effect of providing information about energy consumption. In their pioneering work, McClelland & Cook [MC79] used in-home displays to inform their study participants about the monetary cost of their current electricity use. Over a period of eleven months, the study's participants reduced their consumption by an average of 12%. In a similar setting, [HM+86] provided information about electricity and gas usage that led to savings among participants in two out of three cities. Van Houwelingen & van Raaij [HR89] investigated the influence of feedback frequency on savings and observed stronger effects when feedback was continuously provided instead of on a monthly basis. Brandon & Lewis [BL99] showed that low-use consumers who received feedback information actually increased their energy usage. Schultz *et al.* [SN+07] applied the latter finding outside the ICT context (using door hangers in hotel bathrooms), showing that descriptive normative feedback (i.e., feedback on what other people typically do) leads to an increase in electricity usage among below-average consumers, whereas a combination of descriptive normative and injunctive normative feedback (i.e., feedback on what other people appreciate) does not. Abrahamse *et al.* [AS+07] studied the conservation effect for direct and indirect energy use. They provided feedback to households via an Internet-based tool (one web page for the treatment group and another one for the control group). They implemented a *goal setting* feature and *tailored feedback*. The energy savings were also *framed* differently (in monetary savings and in resource savings). They studied the combination of interventions over five months in 189 households and achieved 5,1% of direct savings, whereas the control group used 0,7% more energy. Knowledge on energy consumption and on energy-saving behavior was also measured before and after the field test. The treatment group revealed an increase in knowledge level about energy conservation and in the adoption of conservation behavior. Loock *et al.* [LS+13] provide a helpful overview (Picture 26) of nine IS-based interventions and outline comparisons and saving effects for various feedback elements. However, there is no clear best practice regarding the usage of psychological nudges. Additional meta-reviews outline a wide range of interventions for electricity usage [AM10][AS+05][Da06][GS+11] [MD+10][NR+09]. More recently, a number of field studies have investigated the effect of real-time consumption information on driving behavior [MW+09][EG+10][GR+10]. The recommendations given in these studies target utilities and policy-makers and only implicitly support systems designers.

With the focus on the feedback interface design, human computer interactions often research innovative consumption or feedback interfaces. As an example, aquariums are chosen to convey intuitively and figuratively consumption information about electricity [CL+12]. The set up is intended to be persuasive and to ensure continuous interaction. The digital aquarium displays a dynamic ecosystem that is a metaphor for the energy consumption (Picture 25). There are 19 levels, each representing a gradually different energy consumption situation. The scenery may degrade or become richer (with more or less divers fish species and plants). It incorporates first of all the information framing. There are no *hard fact* figures of energy consumption. The modifications of the ecosystem affect on an emotional level. Due to a public installation place it is supposed to induce a social comparison or peer-effect. Tests for two variants in two research laboratories revealed different effects. One of them was more successful in saving energy during the eight-week trial than the other. There was actually just a small saving trend after the introduction but savings did not persist on the long run. However, when both persuasive feedback mechanisms were compared to purely textual displays, the digital aquarium always performed better in decreasing energy consumption.

Another similar project, BluePot [Li13], displays a flower as representation for energy consumption. Even though the experiment settings didn't follow an extensive research approach (10 houses, baseline of two weeks, four week treatment time), it successfully conveyed energy consumption.

Picture 25: Digital aquarium as a feedback device

Author(s)	Intervention	System	Resource	Sample	Results
McClelland and Cook (1980)	Feedback	In-home display	Electricity	101 families	Feedback group saves 12%
Hutton et al. (1986)	Feedback, information	In-home display	Gas, electricity	3 cities	4–5% savings in two out of three cities
Van Houwelingen and van Raij (1989)	Feedback (frequency), goal setting	In-home display	Gas	325 families	Continuous feedback + goal setting leads to the highest savings (12.3%)
Dobson and Griffin (1992)	Feedback (breakdown, frequency)	Software	Electricity	100 house-holds	Continuous and appliance specific feedback leads to savings of 12.9%
Abrahamse et al. (2007)	Feedback, goal setting, information	Web portal	Energy	189 customers	Feedback + goal setting + tailored information leads to 5.1% savings
Loock et al. (2011)	Feedback (content)	Web portal	Electricity	220 customers	Injunctive feedback always reduces consumption, descriptive feedback leads to increased consumption for below average consumers
Graham et al. (2011)	Feedback (content)	Web portal	Fuel	128 students	Combination of monetary and environmental feedback works best for reducing car use
Peschiera and Taylor (2012)	Feedback (content), competition	Web portal	Energy	44 dorm rooms	Social feedback is more effective when using peer norms instead of impersonal energy consumption norms
Chen et al. (2012)	Feedback (content)	Web portal	Energy	89 dorm rooms	Social feedback is more effective than individual feedback

Picture 26: Overview of related studies testing IS-based feedback Interventions

2.2.2. Qualitative evidence

With respect to feedback technologies, qualitative empirical research mostly focuses the development of interface prototypes. The research is often qualitative by nature due to the small number of study participants, which does not allow for quantitative analyses. However, if done carefully, the studies allow for improving aspects such as ease of use, understandability, and general acceptance. Selected interfaces from such research endeavors are presented in the following.

Weiss *et al.* [WL+12] developed an application for feedback on electricity consumption via mobile phones called eMeter. The app combines direct and indirect feedback for aggregated consumption data. It also helps to determine the current electricity consumption of individual appliances via an interactive “measurement feature”. Weiss and colleagues integrate simple *goal setting* and *social comparison mechanisms*. For system evaluation, a study with 25 users was conducted. The overall rating of the application was positive. Especially the interactive measurement feature intrigued them.

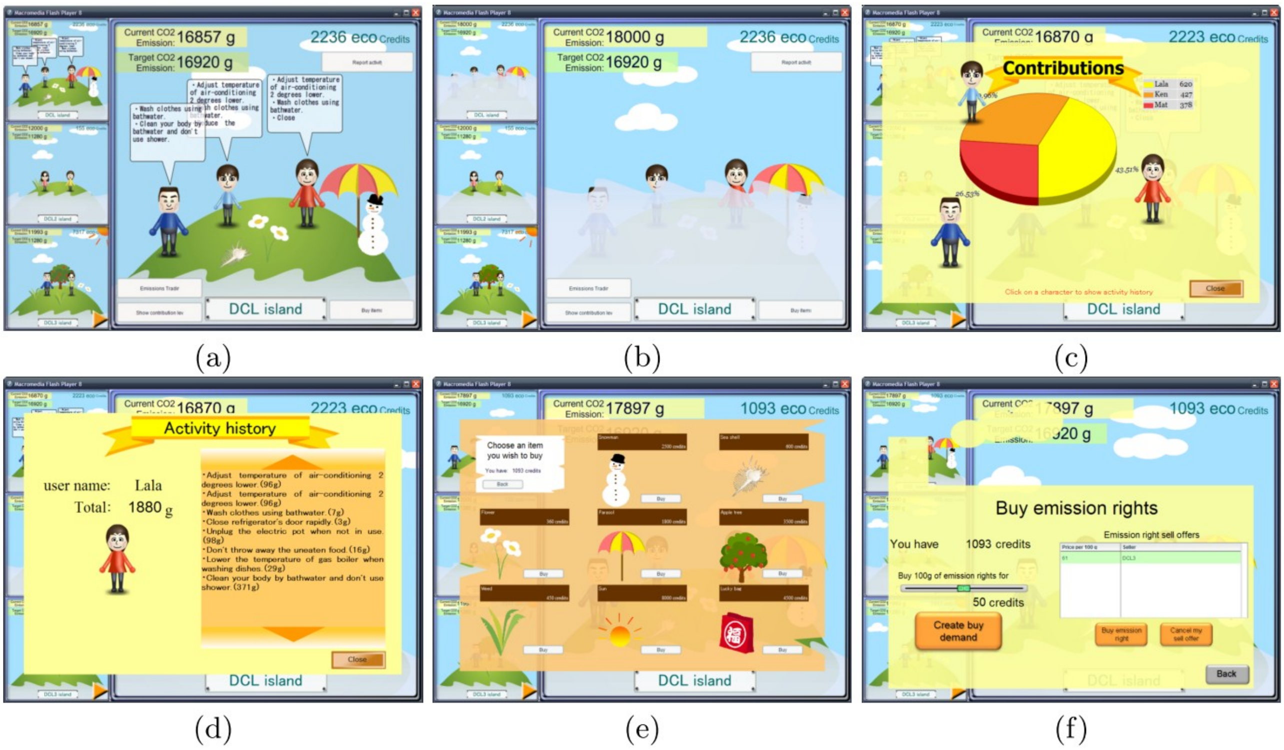
Picture 27: User interface of a mobile feedback device

Another example for a mobile application is the persuasive tool *EnergyLife* studied by Gamberini *et al.* [GJ+11][BJ10]. The particularity of this eco-feedback system targeting electricity consumption is the integration of *gamification* concepts. After having completed energy curtailment tasks, the users enter new levels and are provided with more information about saving measures. Moreover, social networking aspects are included via a gaming community where users can exchange experiences and practical wisdom. User trials showed a high degree of satisfaction with the prototype and strong appreciation of the gaming aspects. However, over time the frequency of usage dropped considerably – despite the very positive evaluation of the app by the users. Gamberini *et al.* explained the phenomenon with the fact that the game became *too complicated* with higher gaming levels. A limitation of the study is the small number of users (N=24).

For yet another electricity portal, Erickson *et al.* [EL+13] investigated the performance of their interface design. They applied consumption feedback, incentives, comparison, and goal setting to encourage saving behavior (Picture 31). The "trend-rank" metrics (1) and the "consumption insight" page (2) include *descriptive feedback* functions. A comparison with historical consumption data was also implemented (3), as was a *goal setting feature* (4) and a page providing saving advice. Erickson *et al.* extensively evaluated their interface qualitatively through user surveys and interviews. Responses about the intensity of usage revealed that people were mainly interested in their baseline consumption and reported frequent visits at the beginning. Only 20% reported regular use. Neither did the respondents pay much attention to the Facebook chat nor to the alerts. Also, the comparison to neighbors did not encourage the respondents very much to act upon the information.

Picture 28: The Dubuque Electricity Portal

Shiraishi *et al.* [SW+09] present EcoIsland (Picture 29) that aims at persuading whole families to engage in energy conservation. The intervention artifact incorporates various persuasion techniques and the context of the intervention is designed as a game. A display is installed in the living room or a central place of the household so everyone can monitor the course of the game (*social facilitation*). The metaphorical goal of the game consists of preventing a virtual island with all household members as habitants from sinking (b). The factual goal was reducing CO_2 emissions by 6%. This can be achieved by pro-environmental behavior (*intrinsic motivation*). Eco-friendly behavior is also directly rewarded in a virtual currency (*extrinsic motivation*). It allows purchasing decoration for the island (e) or trading emissions with other households (f). Activity reports in form of speech bubbles (a), history reports (d), and contribution charts (c) offer all family members information on the progress of the other participants (*social facilitation, normative feedback*). Six volunteer families tested the system during four weeks and reported their saving activities with their own mobile phones. Electricity meters were used to monitor the air heaters. After one week for baseline measurement, the game started. The field experiment was followed by a survey and an analysis of the electricity meter. Even though the households reported an increased consciousness about pro-ecological behavior, no reduction in heat energy was observed. The authors explain this by the short experiment duration. As the survey reveals, the evaluation of the persuasion techniques was generally positive. Users explained that they were principally motivated by extrinsic motivation (rewards in the form of the virtual currency) rather than intrinsic motivation (preventing the island from sinking). The *goal setting* function (default goal of saving 6% CO_2 emissions) was perceived as easy to achieve by most participants. Others suggested an individual goal setting function to avoid frustration when the goal cannot be achieved. Furthermore, social facilitation increased the activity reporting of family members when all the members joined the game in week three. Social comparison and competition captured the user's interest and motivated them to use the system. The authors conclude that easy and short term goals motivate users to change their behavior and suggest a dynamic goal setting function in comparison with further achievements for future designs.



Picture 29: Eco Island [SW+09]

2.3. Water saving interventions and interfaces on the household level

Despite the extensive body of literature on consumer behavior and the effects of consumption feedback in general [OE08], related work on water consumption is sparse. Compared to the number of studies focusing on electricity and gas usage, very few studies exist on the effects of consumption information on water demand. Furthermore, the majority of the existing contributions lack a solid, theory-guided foundation on the underlying behavioral principles. This constitutes both a challenge for the design of water saving interventions and call for action to conduct further research.

Water saving interventions are an important part of water demand management (WDM) strategies of utilities [WS+10]. They consist of water metering, flow restrictors, efficient devices, and educational campaigns. Further information on WDM follows in chapter 4, thus, this section only focuses on describing interventions and interfaces on the household level.

2.3.1. Retrofit and rebate interventions

Early water conservation studies revealed significant influence of conservation campaigns [IJ06][LT+11][MD+04][GE+83]. First rudimentary programs consisted of information campaigns, plumbing retrofit rebates, and measures to restrict flow rates in general. Data from eight urban water agency services in California over seven years revealed savings of up to 25% of water demand [RG00]. Generally, mandatory policies (restrictions and allocations) reduced aggregated water demand more than the voluntary ones.

2.3.2. Feedback interventions

A large part of WDM strategies focuses on awareness. Yet the effect of such measures is still being controversially discussed. Campaigns showed savings in the range of 10-25% [SN+00]. Large comparisons in the U.S. revealed that the effectiveness of feedback also depends on regional aspects. The Western parts of the U.S. showed a more pronounced responsiveness to saving campaigns than Eastern parts, probably because water scarcity is more eminent there [Ni92]. In this context, Beal *et al.* [BS+13] provide interesting insights into perceived and actual residential end-use water consumption: Based on an extensive information gathering (smart meter data and resident inputs), they identified that individuals with self-reported low water consumption tend to actually be among high consumers. Vice versa, individuals with self-reported high consumption tend to consume less than average. Feedback might help to overcome existing misconceptions among alleged low consumers. The next section outlines the different feedback means and interventions for water consumption and analyzes the deployment of psychological effects.

2.3.2.1. Mailings

In early water meter studies [GE+83], interventions mainly consisted of written feedback with daily and weekly summary graphs. The graphs included:

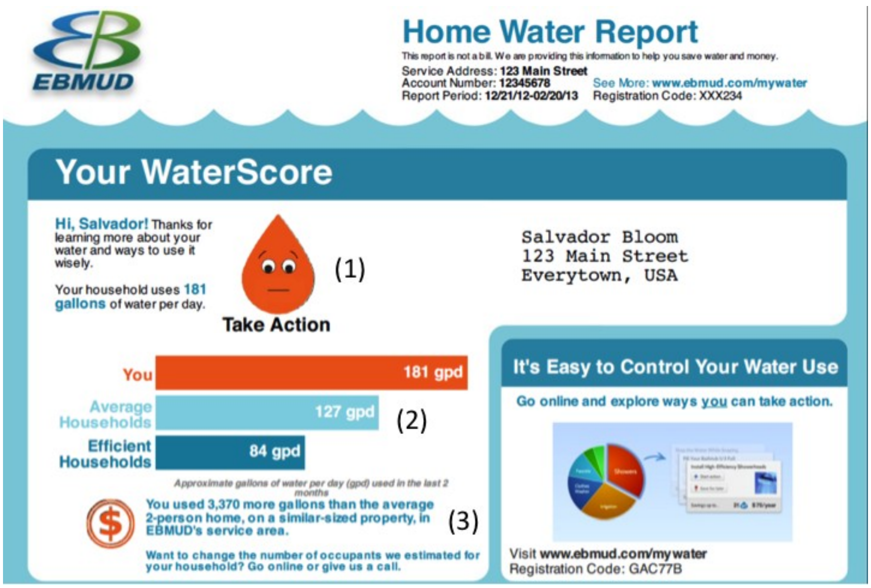
- The total amount of consumed water,
- Charts that plot the consumption over time,
- The percentage of increase or decrease within a certain period of time,
- Comparisons to average demand figures (descriptive normative feedback), and
- Smiling or frowning faces (injunctive normative feedback).

So, besides historical analysis and comparison effects, normative feedback was already utilized in early studies. Smiling and frowning faces are still employed in many interfaces today. However, in the early studies, large water saving effects were only induced by the installation of physical water conservation devices such as flow restrictors. The poor performance of early feedback campaigns might originate from the feedback delay and the high level of aggregation of consumption data (consumption metering at the main water inlet).

Aitken *et al.* published a study on the effects of postcards as feedback mechanism [AM+94]. They explicitly tailored the feedback in a way that it creates cognitive dissonance. Aitken *et al.* compared the weekly average consumption with a similar household and incorporated a reminder of the commitment in the participation agreement in case the consumption was above average. In that way, diverging attitudes and behaviors were used as a means to increase saving effects. However, the conservation effect was rather low with an average of 1,7%.

Fielding *et al.* [FS+13] use a smart metering infrastructure for measuring water consumption every 5 seconds for 221 households in Brisbane, Australia. In addition to a comparison of WDM strategies, the novelty of their research is justified by the long intervention duration of 15 months. This allows assessing long-term effects of various WDM strategies. They studied three treatment groups all receiving four postcards during the intervention with saving tips. The control group did not receive any postcards. Each of the three treatment groups received a special intervention: (1) The *information-only* group received saving advice. (2) The *descriptive norm* group received descriptive normative feedback. The post card conveyed the information that a majority of other similar households were saving more or less water. (3) Alone the *water end-use feedback* group incorporated individualized information on the household's overall water consumption and a breakdown for various consumption activities. Pie charts per fixture and per person were displayed on the postcard. However, such a time-intensive analysis did not allow sending these detailed postcards at all four time points even though software (TraceWizard) was used for the meter data analysis. All in all, there were significant savings for all treatment groups of 11,3 liters per person per day. These savings could not be sustained during the whole intervention. After twelve months the same consumption levels as before the intervention were attuned.

Mitchell *et al.* [MC+13] report about 5% of water consumption savings due to the Home Water Report (Picture 30). 10,000 households in Oakland, USA received paper-based or electronic Home Water Reports periodically during one year. The individualized mailings include social norms (2), injunctive and descriptive feedback (1,2) and personalized comparison effects (3) for visualizing the energy consumption and inducing usage reductions.



Picture 30: Home Water Report [MC+13]

2.3.2.2. Web-based campaigns

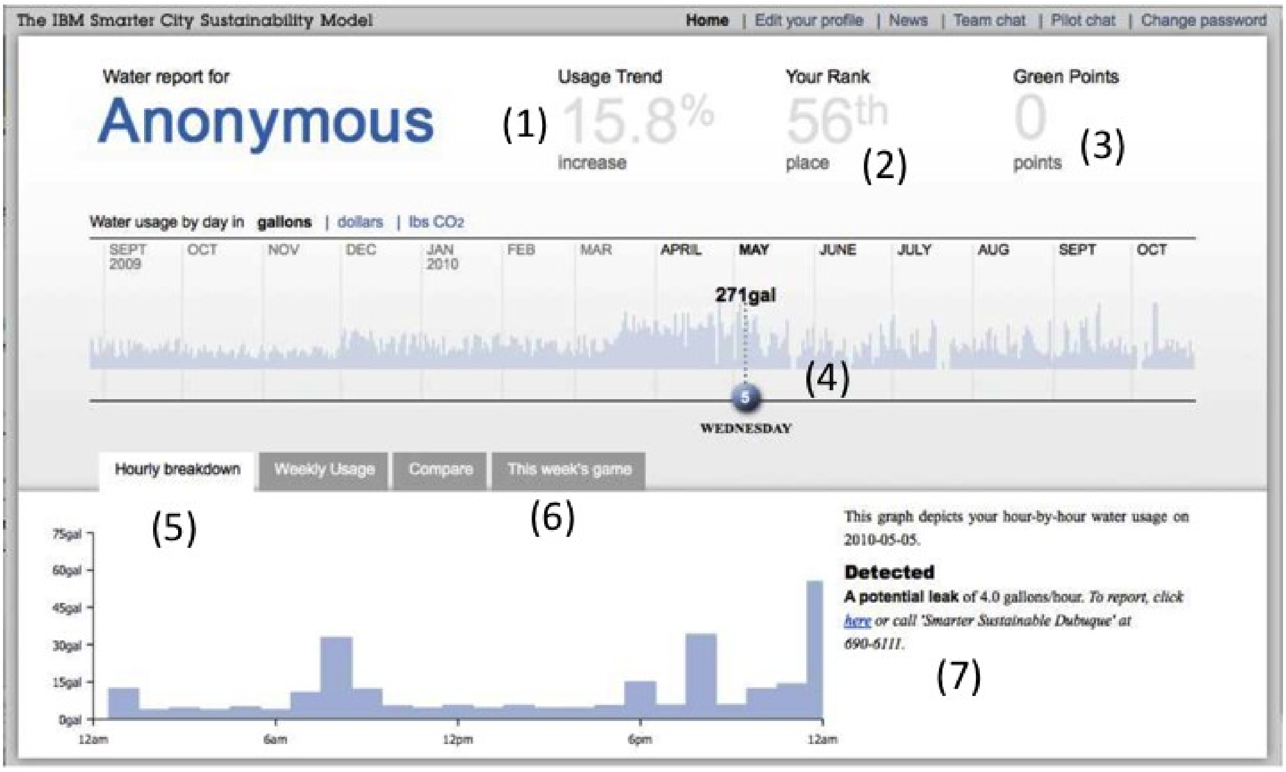
In the course of the increasing Internet usage, web portals gained importance as a means to convey energy-related interventions. Web portals facilitate rich content, frequent updates, virtually infinite interaction mechanisms, and easy integration in social media platforms. Moreover, using web portals, developers can base the selection and instantiation of interventions on rules that allow for a very high degree of individualization.

Various studies leverage that potential, including Mitchell *et al.* [MC+13] and Petersen *et al.* [PS+07]. Petersen *et al.*, for example, delivered consumption feedback for dormitories via the Dormitory Energy web site. The intervention started with the announcement of a competition (on posters and flyers). The web site included an overview of the environmental and economic value of resource reduction (*information framing*) and a comparison of the reductions among dormitories (*social and normative feedback*). The implementation induced savings of 3%.

Along the same lines, the Eco-Pioneer program [We08] included a website and in-home displays that visualized electricity and gas consumption (Picture 21). The interface showed aggregated consumption information and could display 200 character messages. The users could also activate an audible and visible alert function. The webpage also supported an analysis of historical consumption data. Among the 50 study participants, average water savings of 5% have been shown. With a duration of 12 month, the study represents one of the rare long-term observation.

IBM Research conducted a pilot project with 303 smart water meters within the Smarter City Sustainability Model in the city of Dubuque, USA [NL+11]. The meters were connected to the households' main water inlets. In this project, 50% of the households were granted access to an online portal that provided information on water usage, which was based on hourly aggregated data. The web interface contains a personal ranking feature (*social comparison* (2)), information on general saving achievements (1), historical consumption data (4), a leak detection indicator (7), an hourly and daily breakdown of water consumption (5), and *gamification* elements (3,6). IBM reported water savings of 6.6% over a period of nine weeks compared to the other group. Comparable feedback studies for electricity typically lead to savings of 3% to 4%.

IBM's study is valuable in the context of automated billing and also demonstrates the potential of timely consumption information on demand. Yet it does not capture the effects of real-time feedback at the point of use. Furthermore, since it represents the state-of-the-art concerning what can be achieved with current smart water meters, it is apparent that smart meters cannot provide alone the water saving results of other sensing and feedback techniques. Moving beyond the state-of-the-art, real-time feedback systems for water consumption are explored in the next section.



Picture 31: Water Consumption Analysis in IBM trials

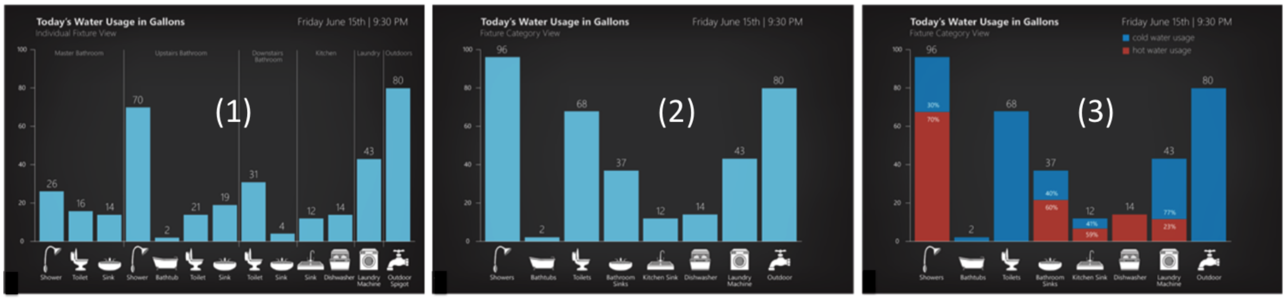
2.3.2.3. In-home displays (IHD)

In their meta-study, Neenan & Robinson pointed out the power of real-time feedback on individual action [NR+09]. For water usage, several prototypes and early products exist that provide this kind of feedback.

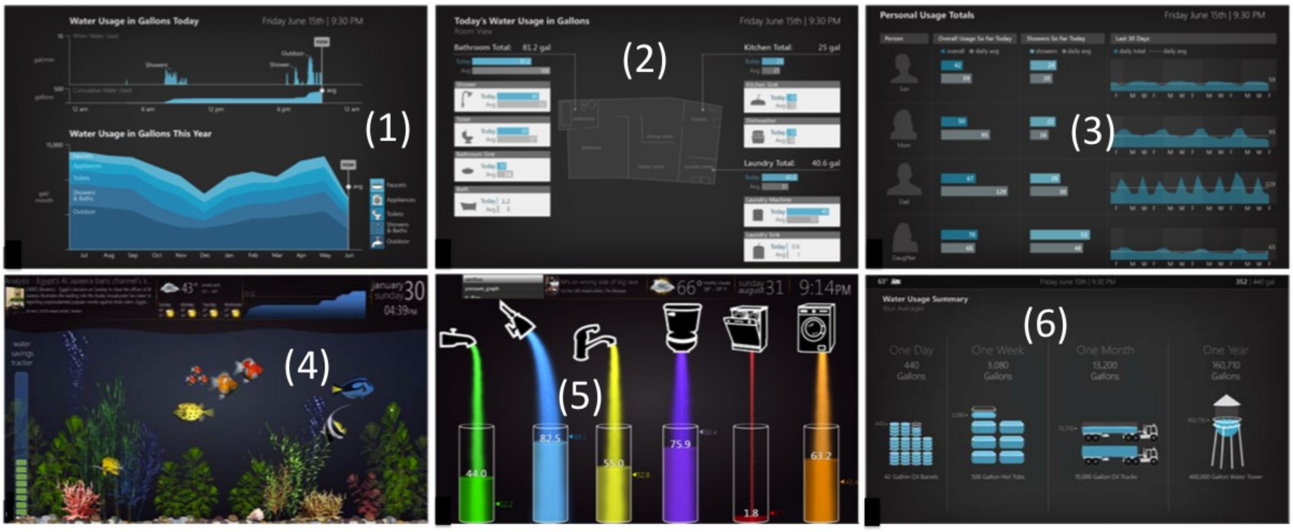
Froehlich *et al.*, for example, designed and evaluated two types of feedback displays that present detailed consumption information [FF+12](Picture 32).

Picture 32: Feedback displays with fixture information [FF+12]

In their study, Froehlich et al. used the first type of display to examine different levels of data aggregation (Picture 33). They created an overview of all fixtures available in a household (1), of fixtures aggregated by type (2), and of the hot and cold water consumption per fixture type (3). The second display was used to test various graphical ways to present the information (Picture 34). The researchers investigated usual daily and yearly water consumption overview in graphs (1), fixture-based breakdown for the yearly consumption, and *social comparison* and *social norm* functions (2). Moreover, they visualized the floor plan of an apartment and indicated the daily and average consumption per room (e.g., for the kitchen, bathroom, and laundry room) and fixture (e.g., for shower, bath, and dishwasher). More intensive *social feedback* included information on individual members of a household (3). *Social facilitation* allowed everyone to follow the progress (a comparison of average usage and daily consumption) of each household member. Froehlich *et al.* also used an aquarium metaphor (*information framing*) for representing the water usage of household members and combined the non-factual visualization with practical information (e.g., date, time, weather, news) (4). Yet another version showed a bar graph that additionally visualized the volume by means of a water jet of different width (5). Finally, the last design probe indicated daily, weekly, monthly, and yearly consumption in the corresponding number of jugs, baths, trucks, or water towers. Their study concluded that users prefer individual fixture information, breakdown of water in hot/cold, and consumption indicated both in liters and money. The vast majority of users regarded comparisons of their recent domestic consumption with their past consumption as important, and attributed high value to specific goals. The work of Froehlich *et al.* provides an excellent starting point for the design of future feedback displays. As the current version does not include real-time information, the integration of real-time feedback can be seen as the next step towards powerful water conservation products. Moreover, future studies on interfaces should go beyond relying on user ratings and include measurements of water consumption in order to judge upon designs.



Picture 33: First display types [FF+12]



Picture 34: Second display types [FF+12]

2.4. Water-saving interventions and interfaces on fixture-level feedback

In-home displays that report aggregated water consumption have been shown to produce noteworthy saving effects. However, they most likely do not reap the full potential of feedback interventions, as the information neither is put forth at the point of consumption nor at the moment the decision on water usage is made. This section explores interventions and interfaces that provide feedback on the faucet level, mostly even in real-time and in a way that a user can receive the feedback message when she is still in the process of taking water from the tap.

2.4.1. Displays and monitors

The in-situ feedback device amphiro a1 (Picture 39) is built to reduce water and energy consumption in the shower. The display indicates information on three main areas (see Picture 39). On the top (1) and in the middle (2) of the display, alternating metrics are shown, including the water temperature, an energy efficiency scale from A to G, and the water consumption of the ongoing shower in liters. The energy efficiency scale (*normative feedback*) depends on the length of the shower and the water temperature. After the shower, the middle section toggles between water volume and energy used in watt-hours or kilowatt-hours. The bottom part (3) visualizes a climate animation (*metaphoric*) that is selected based on the current level of the aforementioned energy efficiency scale. It is discussed more detailed in section 2.4.3 as persuasive feedback element. In combination with the device, there exists a free online portal where clients can enter an eight-digit code. It is shown at the end of each shower. By entering the code online[3], diagrams and scales display consumption behavior over time.

In Switzerland, a large field trial was recently performed for a period of 5 months with the predecessor of amphiro a1, documenting the benefits and impact regarding hot water consumption [SF11]. Volunteers participating in the trial in typical hand showers performed installation themselves. In total 200 smart water meters were handed out to staff of the public sector; participants installed 49% of the devices, resulting to 3,164 recorded showers in 60 households, from 160 individual users. With feedback information from the display, hot water consumption declined by 22%, (6,400 liters of drinking water) and 210 kWh of energy was saved [TS+12]. In addition, the decline resulted from reduced shower time and not from a change in user habits. The study resulted in several important findings. First, saving effects clearly exceeded those achieved by smart meters for electricity. Second, savings remained stable over time. Third, participants were willing to pay roughly 39 Euros per meter, which would amortize the costs over a nine-month period.

In more recent research projects, amphiro a1 has proven the potential to motivate on average 440 kWh (kilowatt-hours) of thermal energy as well as 8,500 liters of drinking and wastewater per year[4]. The study in cooperation with the Swiss Federal Office of Energy involved 700 households. Beyond the effect sizes, the study confirmed the easy installation of the device, the durability of the product, the participant's acceptance and a strong wish to continue device usage among the majority of the participants. Some weak points such as water entering the display could be improved in the following production lines [TT+13].

Wilis *et al.* [WS+10][SW+13] investigated the effects of the shower monitor WaiTEK (Picture 10) that they installed and operated in 44 households in the Melbourne region. The devices measured baseline consumption within the first month (measurement without display) and thereafter automatically switched into feedback mode to display water consumption. In feedback mode, the devices also provided an acoustic alarm signal when a user-adjustable volume was exceeded. In their study, the authors report an average saving of 27%. Qualitative data reveals that almost 90% of the respondents were satisfied with the monitor. They stopped showering when the acoustic signal appeared and checked on the monitor for the shower duration. The effects mainly resulted from reduced shower duration.

Laschke *et al.* [LH+11] proposed the Shower Calendar (Picture 38), a novel design concept for displaying shower water consumption. Their installation included a projected image of a calendar visualizing water consumption per day and per family member (*social comparison/facilitation*) on the shower enclosures. Colored dots expressed the water consumption relative to baseline water consumption. The device was tested with two families (6 individuals). The dots represented a certain amount of water that could be used for a shower. During water usage the dot was shrinking and the user motivated to stop the dot from fading away (*goal setting and loss aversion*). This work also resulted in soliciting sustainable behavioral changes and water conservation.

The Invetech system (Picture 13) from Simon [Si06] visualizes water consumption and integrates audible reminders for increasing behavioral change. During a six-month trial with 20 units in Melbourne 15% of savings were achieved. The installation is supposed to be feasible with a plumber. However, in comparison to amphiro a1, it is much more complex. The shower senses the start of the shower by the commencement of flow. It displays the flow rate and the shower volume. Based on the average shower usage the Invetech display calculates a consumption target. At the

[3] http://amphiro.com/portal/

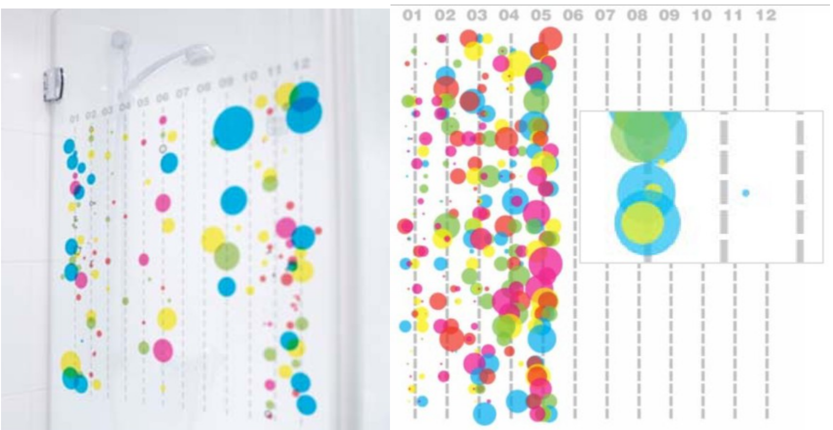
[4] http://amphiro.com/products/a1/

moment the target is reached, audible and visual alarms prompt (*normative feedback, goals*). Besides, the remaining time until the alarm starts is displayed, too. Based on water temperature, the shower meter is also able to determine the energy used for each shower.

Picture 35: WaterBot [AB+05]

Picture 36: The iCat [HM10]

Picture 37: Disappearing Pattern Tiles [BG06]



Picture 38: The shower calendar

Picture 39: Amphiro a1 feedback details

2.4.2. Ambient and persuasive feedback

Arroyo *et al*. [AB+05] tested the WaterBot (Picture 7) that provides real-time feedback at the tap-level. It urges people to turn off the tap when water is not used, for example while brushing teeth or washing dishes. The motivation is supposed to be non-obtrusive. The actual feedback varies depending on the time the tap has been opened. Color changes and an audio effect are supposed to be entertaining (*gamification*). Positive auditory messages reward the user for closing the tap. Additionally, a form of social comparison is incorporated. Two green bars indicate the water consumption of oneself and of a peer. Although there has been no systematic experiment conducted to quantify the water savings from the device, observations and user reports suggested a behavioral change that could reduce water consumption. Users easily understood some feedback functions even without knowing what the device was for.

The WaiTEK shower monitor [WS+10][SW+13] also includes acoustic feedback. A beep starts after exceeding a certain amount of time. Users felt influenced by the signal, however, they admitted to getting used to the one-minute beeping and continued with the shower.

Kappel & Grechenig [KG09] developed a shower water meter (Show-me) that visualizes the amount of water used during one shower in the form of LEDs assembled on a stick, and installed the device in several households in Austria (Picture 8). The results showed a decrease of the average shower water consumption of approximately 10 liters. This suggested promising water saving potential in the shower with regards to using visual displays for delivering feedback.

Kuzentsov & Pavlos [KP10] proposed UpStream (Picture 9), a system combining low-cost water flow sensing coupled with persuasive displays. A microphone connected to a low cost Arduino micro-controller measured water flow. The results showed an overall decrease in water consumption and also raised concerns regarding gender-induced behavioral changes. However, their work was limited in its technical scope and explored only short-term effects.

Amphiro a1 also incorporates a persuasive feedback function, with a polar bear standing on an ice floe. The natural habitat of the polar bear and the animal itself disappears step-by-step depending on the duration of the shower and the temperature of the water. All the components represent an animation that shows directly the effects of showering in an energy efficiency context.

Following another approach, the iCat [HM10] represents a rather rarely studied feedback approach: *social feedback*. A robotic agent operated in simulation experiments gives feedback about the selection of a corresponding washing program. The iCat resembles a cat's head and integrates moving lips, eyes, eyelashes, eyebrows, and voice recordings. Ham & Midden [HM10] revealed during their experiments that negative social feedback was more effective than positive one. Additionally, in comparison with factual feedback, the social feedback showed stronger saving effects.

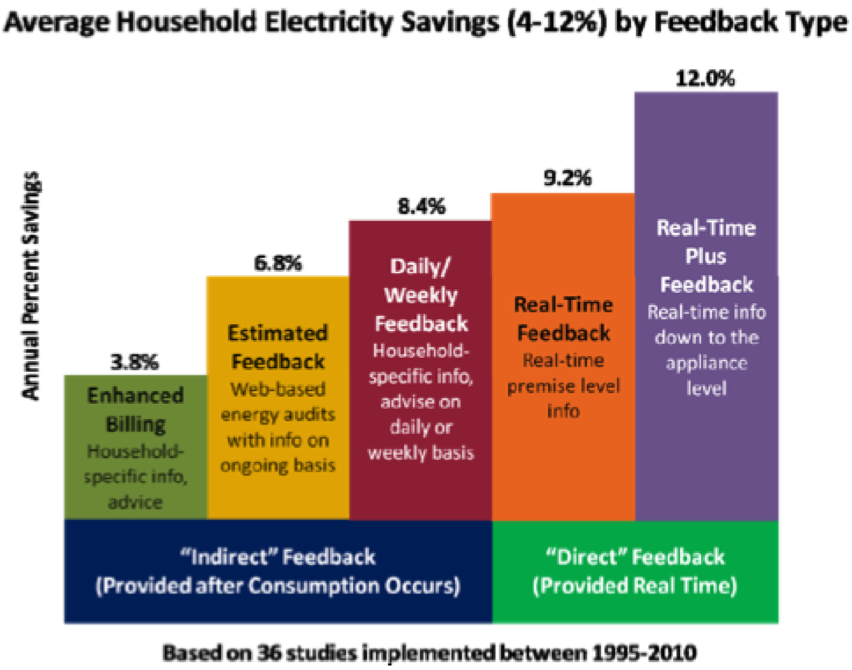
Finally, Backlund *et al.* [BG06] designed bathroom tiles painted with a floral décor made out of thermo-chromic ink. It reacts to heat and fades away during long and hot showers. It is intended to immediately display the consequences of extensive hot water usage. As to date, the tiles have not been used in any study yet, and no empirical data or experience is available.

2.5. Lessons learned from electricity

This section outlines the results of several meta-studies on electricity usage [AM10][AS+05][Da06][GS+11][MD+10][NR+09]. It summarizes important details concerning effect sizes, the study design, and the psychological concepts used in electricity metering that might be transferred to water-centric interventions.

2.5.1. Effect size

The graph (Picture 40) is taken from a meta-study [MD+10] on electricity conservation campaigns that deployed different forms of feedback. It shows saving effects in percent by feedback type and reveals a general pattern: The "closer" the feedback information is to the actual consumption behavior (both in terms of time and space), the higher are the saving effect [NR+09]. Effect sizes range from below three to twelve percent, depending on when, where, and at what level the feedback is given. For DAIAD, this highlights the importance of implementing real-time feedback provided at the place where the consumption decisions are made.



Picture 40: Overview of saving effects for different feedback types [MD+10]

2.5.2. Study design

Allcott & Mullainathan [AM10] suggest to systematically structure interventions after the following aspects:

- In-situ testing
- Randomized controlled field trials in a representative population
- Randomizing (letter) content across groups
- Program marketing and program evaluation
- A phased implementation (especially when technology is involved [MD+10])
- Not only motivated people should have the possibility to opt-in
- Participation is higher, when an *opt-out* possibility is given
- Combine other methods such as interviews and laboratory experiments

The focus of the program should be clearly communicated to the participants. Load shifting programs have been proven to be generally successful. Interventions referring to reduce energy during specific time periods have been shown to be more effective than general energy efficiency campaigns. Erharhdt-Martinez *et al.* [MD+10] found a correlation between study duration and energy savings. Savings tend to be lower for longer studies than for shorter ones. This might be due to the influence of the researchers during short-term studies (perhaps due to more questionnaires per time) and/or due to the interventions loosing power over the course of the experiment. Energy savings also depend of era and region [MD+10][Ni92]. Crises directly affect changes in culture, politics, and lifestyles. For regions being familiar with drought, energy savings are higher. Households with higher baseline consumption typically show larger saving effects because the scope of reductions is larger. Furthermore, the usage of the same energy display device in different countries (same sampling protocols and population frames) led to strikingly differing conservation effects from 2-18% [NR+09]. Households that are already efficient showed less saving potential [MD+10][Da06]. Therefore, it is vital to survey/measure baseline consumption data, preintervention behaviors and household characteristics. Overall, most of the studies – especially the earlier ones and the ones with a technical focus - reveal a poor study design (small sample, short duration) [MD+10].

2.5.3. Feedback type

A combination of direct and indirect feedback (enhanced billing and energy display devices) is especially recommended for effect persistence [MD+10]. In-situ electronic visual monitors or acoustic alarms provide quicker and more frequent feedback than or web-based portals [WS+10][GS+11][AM10]. ICT also fulfills the requirement of offering tailored feedback much easier [AS+05][NR+09]. For further information on different types of feedback, see section 4.2.2.3.

ICT-based feedback devices require energy on their own. While the saving effects will most likely justify the devices' electricity consumption, some practical issues of power supply are more important: As mentioned by Weiss *et al.* [WL+12], residents might not be motivated enough to change batteries when they are empty. This leads to devices that are not used long enough to justify the resources that had to be invested for their production. Moreover, especially for wet environments, connections to the mains are often not feasible for at least two reasons: Placing wires constitutes an additional cost factor, and plumbers are often not willing to or allowed to performing such tasks. Additionally, complex installations and the need for electricians decrease the willingness of consumers to have such systems installed. Therefore, self-powered devices are often the technology of choice. For future DAIAD designs, these issues should be kept in mind when it comes to power supply.

2.5.4. Negative effects

Some additional effects and theories should be taken into consideration in order to reap the full benefits of feedback.

- The *human exception paradigm* indicates that humans are above nature; therefore, they sometimes feel that they do not have to regard nature as they consume resources [SW+13]. It should be possible to overcome such attitudes by targeted awareness campaigns that use the concepts of cognitive dissonance.
- The *spillover effect* explains that accordingly to the attitude-consistency concept, people tend to behave in line with their attitude. So, once the pro-environmental behavior is seeded, it might jump over to other areas of life.
- Opposed to the spillover effect, the *rebound effect* and *moral licensing* represent an obstacle for future campaigns [TS+13]. In contrast to a positive dissemination of pro-environmental behaviors, individuals might feel like they are entitled to spend the conserved energy/resources in another area of life. For example, people may increase their water usage after installing toilet dams and faucet aerators because they think they would automatically save enough water anyway.

3. Big Water Data Management and Analysis

In DAIAD we will develop water sensing technologies providing real-time and highly granular data regarding water consumption in domestic environments. This level of detail is unprecedented in the water domain and introduces technical challenges and opportunities not currently addressed in the management and analysis of water consumption data. The increase in data volume and velocity DAIAD will provide, natively establishes water consumption data as a Big Data source.

Big Data, a term coined to describe novel technologies and challenges as a result of the massive, interconnected and highly valuable data produced from modern ICT systems, is already relevant for a number of business areas and scientific domains. Novel, cost-efficient and highly scalable means to manage and analyze data, result in extracting actionable insights and value from data at an unprecedented scale. DAIAD is the first FP7 project actively treating water consumption data as a Big Data source, a novelty which requires an adaptation of relevant technologies for water data, as well as development of methods to accommodate the specific needs of water stakeholders.

3.1. The age of Big Data

ICT has enabled the generation, storage and processing of unprecedented volumes of data, resulting in a significant transformation of Data, Information, and Knowledge Management. Indeed, as any consumer of information services has also become a producer of information, data is generated and stored at amounts and forms never seen before. At the same time, we increasingly rely on analysis and knowledge obtained from the data to innovate and advance in science and technology.

Our expectations from Information Systems have greatly evolved, with relational databases and data warehouses not dominating the landscape. Today, Information Systems must be scalable and extendable to cope with growing amounts of data, scaled over local or wide area networks at the scale of one or multiple datacenters. They must support offline and also online analysis of data as it becomes available, as well as advanced analysis to extract actionable knowledge from overwhelming volumes of data. Moreover, new architectures emerge in order to exploit the recent developments in software and hardware design.

Factors that have underpinned this shift are the generation of unprecedented amounts of data as a direct product (e.g. satellite imagery) or by-product of commercial applications (e.g. web purchases, social networks), the availability of open data on the Web, the increasing popularity of online social networks and user-generated content, and the emergence of the Internet of Things (IoT), where devices equipped with sensors provide constant streams of data.

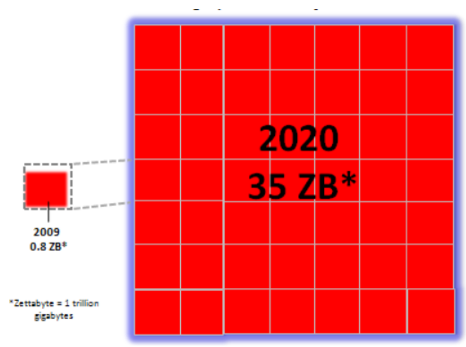


Figure 1: The Digital Universe 2009-2020: growing by a factor of 44; http://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf

In various fields and domains, the **Big Data revolution** landmarks new business opportunities and operation models, fueled by knowledge and value extracted from unprecedented levels data analysis. A new generation of specialized workers (*the data scientists*), apply data analysis technologies to identify correlations and extract insights in vast amounts of highly complex data. In some cases the analysis results enable a company to offer better services, by recommending products to users, or by streamlining its supply chain. In other cases, data of seemingly no value (by-products, known as *exhaust data*) can give birth to a completely new market activity (e.g. the analysis of mobile logs to extract time-travel maps for navigation).

Several studies have identified Big Data as the next step for **financial growth**, establishing it as a leading R&D activity, and an enabler for **competitiveness** in most economy fields. This is the case of the EU as well, both as a whole and in its member-states. Businesses of all sizes already apply Big Data technologies and reap the benefits of leading this innovation. However, controversy is also present. An ongoing criticism towards Big Data advocates is that they ignore the time-honored and tested core scientific methodology. Instead of extracting conclusions based on observations, assumptions, modeling and statistically/experimentally testing their hypothesis, Big Data practitioners at large rely on data alone.

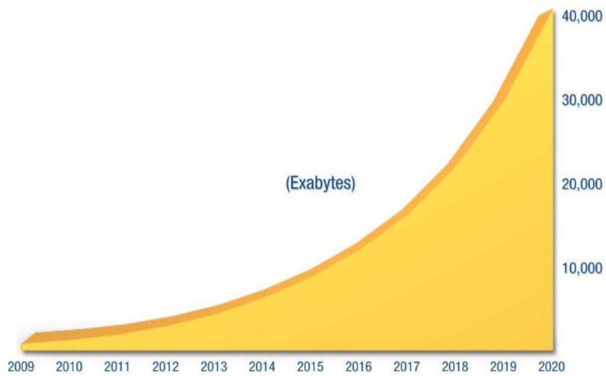


Figure 2: The Digital Universe, 50-fold growth from the beginning of 2012 to the end of 2020 - http://idcdocserv.com/1414

On a technical level, Big Data has given birth to entirely new paradigms and concepts for data management and knowledge extraction. The traditional relational databases and data warehouses could simply not handle the scale and complexity of the new data in terms of size and responsiveness. Further, their inherent design principles (e.g. ACID properties) were in some cases void, as they added overhead where the functionality was not needed. The concurrent emergence of cloud computing, opportunities for massive parallelization on low cost hardware, and the changing storage hierarchy (due to technology advancements and cost reductions) introduced additional opportunities. So instead of relying on '*one-size fits all approach*' regarding data management, new paradigms sprouted to handle specialized, but extremely common and significant problems. Google's Map-Reduce is one relevant example, as graph databases, RDF stores, raster databases, key-value stores.

In the Water domain, the Big Data revolution has not yet arrived. This is mostly due to the size and complexity of the available data. For the vast majority of water utilities, water consumption data remain highly aggregated with a very low resolution in terms of when and where water was consumed. The typical data flow from mechanical water meters means that every few months a simple number (total consumption over a period) is added in a database. Even for smart water meters, the data that is actually stored and analyzed do not reach a critical mass, and thus do not challenge the relational DBMSs in use. Further, this highly aggregated data means that attempting to combine it with other data sources of great quality (e.g. geography, census, weather) will at best produce low quality knowledge, and at worst be in vain.

DAIAD will develop new water monitoring technologies that provide the **missing data**. However, extracting actionable knowledge and insights from them, demands the application of Big Data management technologies. Further, we need to research and develop new technologies that address both the intricacies of highly-granular water consumption data, and its purpose of use. On one hand we have consumers, which must be empowered with actual knowledge and stimuli from massive amounts of data they produce. On the other, we have water stakeholders that want to turn this data in water demand strategies, pricing policies, and even new business models.

In the following sections we provide the state of the art in Big Data technologies, present the dominant paradigms and frameworks, and give an introduction to the software we will extend in the project.

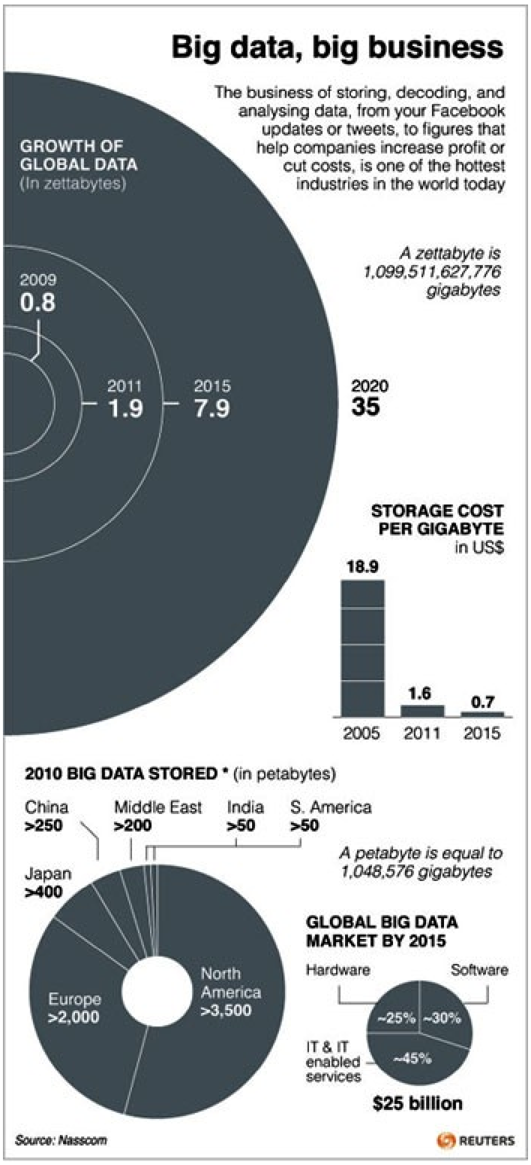
3.1.1. Definition: the Vs

While the database community has always been focused on problems arising from by the size and complexity of data (e.g. *Very Large* DataBases – VLDB - is one of the most prominent conferences in this area), the torrential increase of data and its anticipated further growth, has given a new focus in the exploration of very big databases. In 2001, analyst Doug Laney presented the data growth challenges of data in his report "*3D Data Management: Controlling Data Volume, Velocity and Variety*"[5]. Laney categorizes the challenges in three dimensions:

- **Volume**, which refers to the vast amounts of data generated by humans, the Web, sensors, ICT systems, and all of our interconnected devices. If we add all the data created from humanity up to 2008, their total size is now generated every minute. The volume of the data introduces challenges in data storage, management, and analysis.
- **Velocity**, which refers to the speed at which data are produced and are transmitted. The NYSE captures and transmits 1TB of trade information in each session, while by 2016 there will be 18.9b network connections transmitting data among people and systems. Speed introduces aspect of time-sensitivity, as data must be processed with very strict time-constraints.
- **Variety**, which refers to the different all the different types of data we now generate and use. Structured data, which can be easily handled by relational DBMSs is now the minority. More than 80% of data are now unstructured (e.g. text, video, audio). This variety introduces challenges in the actual integration and analysis of data, such as the meaningful knowledge extraction from social media interactions.

[5] http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

This analysis has been established as a de facto model of Big Data, also known as the '**3Vs**'. In the following years, a number of new dimensions (new Vs) have been added:



- **Veracity**, which refers to the accuracy and trustworthiness of the data. User generated content and social media interactions are a great source potential of information, but their validity and quality is at best questionable. Even authoritative public data sources, as they are becoming openly available, are prone to errors. Veracity introduces challenges regarding the quality assurance from the data (*garbage-in, garbage-out*) and deriving trust in decisions based on inaccurate data.
- **Value**, which refers to the actual value Big Data have. Simply collecting vast amounts of data in system is an exercise in vain, unless these can be leveraged to produce value. Value can be measured as new data, insights, increases in productivity, or even better governance and transparency for the public sector. Extracting value from data introduces challenges in their analysis, not on technical terms, but on the expert-user level. Data scientists, individuals equipped with the technical knowledge to handle data and business understanding to exploit them, are valuable human assets for extracting value from data.

The '5Vs' help the industry and academia to understand the challenges of Big Data and their scope. Based on these, a number of definitions are available in the literature. According to Gartner http://www.gartner.com/it-glossary/big-data/:

> ***Big data*** *is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.*

3.1.1.1. Examples

In the following, we provide a few examples of Big Data and their real-world applications in various business and scientific domains. These only serve as an indication of the wide reach Big Data have in almost all financial and social activities and the tangible benefits they provide[6]. Further, it is important to mention that the reported data size increase daily, hence they definitely exceed what is reported below.

- 40 Zetabytes of data will be created by 2020 (300 times more than 2005).
- 6 billion people (out of 7 billion) will have a mobile phone by 2020.
- Most companies in the USA have at least 100 Terabyte of data stored.

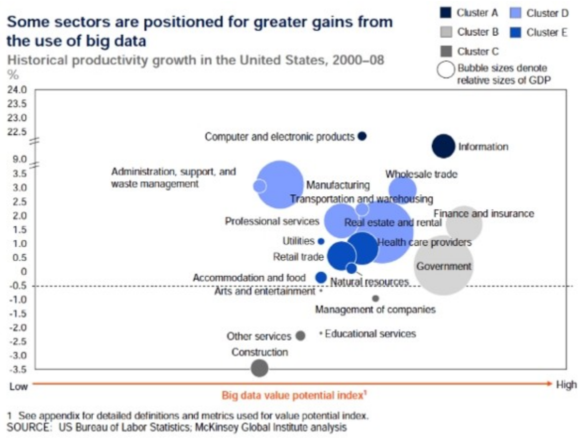
[6] http://www.ibmbigdatahub.com/infographic/four-vs-big-data

- As of 2011, the global size of data in healthcare is estimated to be 150 Exabytes.
- Facebook stores, accesses and analyzes more than 30 Petabytes of user generated data, while its users share more than 30b pieces of content every day.
- Twitter users produce more than 400M tweets every day.
- Macy's, the USA-based retailer, adjusts pricing in real-time for 73M items, based on demand and inventory.
- The human genome can now be decoded in 1 week, instead of 10 years.
- More than 420M wireless and wearable health monitors will be used by 2014.
- Amazon manages more than 100 Petabyte of data for search and consumer recommendations.
- Wal-Mart, the USA-based retailer, provides advanced search facilities to online customers, increasing the number of customers completing a purchase by 10-15%.
- The Los Angeles and Santa-Cruz PD in USA apply predictive analytics for crimes, leading to 33% reduction in burglaries and 21% reduction in violent crimes.

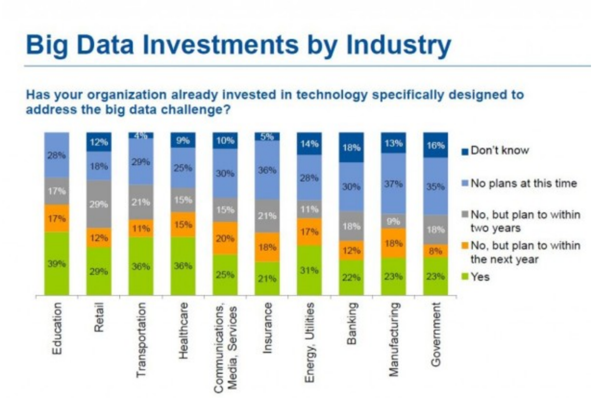
3.1.2. Big Data economy

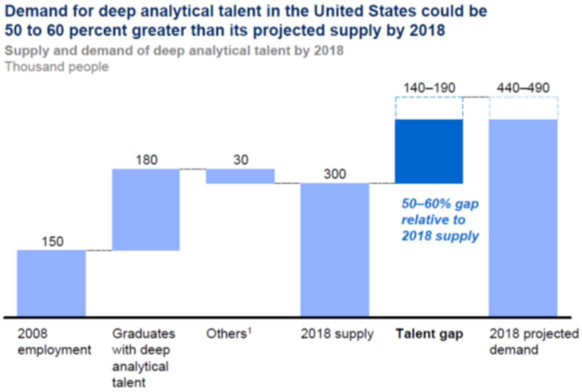
The economic impact of the newfound Big Data economy is recognized to play an important role in enhancing the competitiveness of the private sector and the efficiency of the public sector. For an excellent overview of the business opportunities and value, the reader is encouraged to consult the McKinsey study "*Big data: the next frontier for innovation, competition and productivity*". In the following, we highlight and discuss select insights from the study.

- **Cross-sectoral value**. Big Data technologies have a cross-sectoral reach, by increasing productivity and revenues, while reducing costs. For the EU public sector alone, their value is estimated at $250M per year, contributing 0.5% of annual productivity growth. However, not all economy sectors are positioned to gain from Big Data in the same manner. The information sector is expected to reap greatest growth, as expected. However, utilities (*dominated by energy stakeholders*) will also enjoy a benefit, as will waste management.



- **Readiness**. The realization of potential benefits is dependant from actual investments in Big Data technologies, both in infrastructure (software, hardware), and in human capital. Again, economy sectors exhibit different readiness levels in terms of ongoing investments. Energy and utilities are among the leaders in this respect, along with education, healthcare and transport. However, the human resources required to make sense of Big Data and extract value (*data scientists, business analysts*), are expected to be in shortage. In the USA, a 50% *talent gap* in supply is anticipated by 2018.





As already mentioned, the energy sector is one of the leaders in Big Data uptake, already enjoying financial gains. This is a result of its market size (*available funds for investments and R&D*), wide societal reach (*minor improvements have significant impact*), critical positioning (for security and sustainability), ample investments from the public sector (*R&D, infrastructure investments*), data availability (*smart meters, production data*) and overall ICT maturity. Compared to the water sector, energy stakeholders have always enjoyed more accurate, detailed, real-time data. Further, the distributed nature of the power grid and the currently interchangeable roles of producers/consumers (*renewable energy*) have introduced energy stakeholders in novel ICT systems and paradigms for quite some time. As a result, energy stakeholders already apply Big Data technologies for consumption management and prevention of shortages/outages by combining and analyzing vast data.

In comparison, the water sector is severely lacking. However, as smart water meters and water monitoring technologies are gradually deployed, the availability of data will commence the application of Big Data technologies, and similar benefits. According to a study by Frost & Sullivan [Bu11], the EU market for real time updates on water use, will enjoy double digit growth over the next 10-15 years, with an estimated value of **16.4 billion** by 2020 with the majority of the revenue coming from *data analysis services*. This is an indication of Big Data can have in the water sector, and one of the motivating factors of our work in the project.

3.1.3. Contributing factors

Big Data exists in an evolving landscape of technological, societal, and policy advancements that directly (*or indirectly*) influence the scope and impact of Big Data technologies. These contributing factors are *Open Data, Linked Data*, the *Internet of Things (IoT)*, and *EU's Data Economy*. It is important to understand their impact, in terms of the 5V model presented earlier:

- **Open Data**, i.e. data that can be freely used and reused, are mainly provided by the public sector (e.g. PSI Directive in EU, Copernicus satellite products) and crowdsourcing activities (e.g. OpenStreetMap). In most cases this data have been previously unattainable or available at a prohibitive cost. Therefore they pose a great opportunity for the private sector, and especially SMEs, which have access to data at zero cost. However, big corporations (e.g. OpenStreetMap in Apple Maps) also benefit by expanding the coverage of their offerings at a low price point. Open Data contribute to the volume of Big Data, variety (mostly through crowdsourced information), veracity (access to both high quality authoritative data, but also inaccurate user generated information), and value.

- **Linked Data**, i.e. data that are interlinked and managed with Semantic Web technologies (*Web of Data*), provide great opportunities for facilitating the *fusion* of different data sources, as well as intelligent *reasoning*. While Linked Data is essentially transformed versions of existing data (e.g. DBPedia is the linked version of Wikipedia), they do contribute in the Big Data volume, as they must exist in separate representations. Further, they increase the variety (e.g. RDF), and value.
- **IoT**, i.e. the assembly of interconnected, sensing, identifiable devices over the Internet, has been a vision of the ICT community for decades. The recent advancements in connectivity (e.g. wireless, IPv6), proliferation of low cost sensing devices (e.g. mobile phones), and rise in commercial products (e.g. home automation, personal well-being), has truly materialized IoT (also known as the *Internet of Everything*). It is estimated that by 2020 more than 30b of devices will be wirelessly connected to the IoT[7]. As such, the impact on Big Data is important in terms of volume, velocity and veracity.
- **EU's Data Economy**. The Digital Agenda 2020 and Horizon 2020 have established the Data Economy as the next strategic goal for EU's sustainable growth. The Data Economy is a new form of economic activity founded on Big, Open, and Linked data, and combined will generate value at all stages of the data value chain[8]. Open Data alone will reportedly lead to economic benefits of up to € 40 billion a year in the EU[9] and help unlock $3-5 trillion in 7 sectors[10]. Moreover, the Data Economy can act as a catalyst for solutions in a number of societal challenges and empower all aspects of government and industry as an integral component of ICT.

In DAIAD, we will be applying Open Data as high quality data sources, and IoT technologies. In the following we present the Open Data landscape in more detail.

3.1.3.1. Open Data

Open data and information, i.e. data and information that can be *freely used, reused and redistributed*, is not a current trend, but rather a historically established scientific practice. The goal for providing open knowledge is not only noble, but a necessity for scientific advancement; to share and democratize knowledge, objectively evaluate research, educate, promote cross-scientific activities. In the past 5 years however, open knowledge practices have spread beyond the S&T realm and into the mainstream political and technological agenda. A new movement has been formed, the *open-data movement*.

The reason for this development is the realization of the tangible benefits open data provide. On a *political level*, open data is a facilitator for transparency, accountability and participation, i.e. core democratic values. Access to open data can empower citizens and NGOs to collectively question, audit, and participate in policy making. On the *public administration level*, open data can provide cost reductions through increased technical and semantic interoperability, immediate access to required information for better decision making, and ultimately more efficient services for citizens and businesses. On an *economical level*, open data can boost economical growth, enabling the

---

[7] https://www.abiresearch.com/press/more-than-30-billion-devices-will-wirelessly-conne

[8] https://ec.europa.eu/digital-agenda/en/making-big-data-work-europe-0

[9] http://ec.europa.eu/digital-agenda/en/open-data-0

[10] McKinsey: http://bit.ly/Hv9OgH

private sector to provide products and services of lower cost and increased quality, foster innovation, deliver value added services, and establish an emerging economic activity, identified as the *data economy*.

Open data and information provided by the public sector constitute a significant opportunity for transparency, accountability, better governance, and citizen participation. Reuse of open data can also serve as an instrument for growth, leading to innovation through research, better products and services, new jobs and economic advancement. The Vickery[11] report estimates the benefits of extended PSI (Public Sector Information) reuse for the EU27 economy at 140b€/year (1.7% GDP 2008). These financial benefits will be materialized through the establishment of a *data economy*, led by SMEs providing added value services by repurposing and extending open public data.

Consequently, a number of policies have been implemented in the EU to promote open data. Directive 2003/98/EC on the reuse of Public Sector Information (PSI), and its recent revision proposal to further open up the market for services based on PSI, establish a common binding legislative framework in EU member states for open data & information. Directive 2007/2/EC (INSPIRE) establishes a common legislative, organizational and technical framework to promote interoperable Spatial Data Infrastructures (SDIs) across member states and common sharing of environmental data. Further, the Digital Agenda for Europe and the EU 2020 Strategy imprint a pan-European strategic goal on open data as a facilitator for economical growth, innovation and better governance. Accordingly, several technical measures have been established to promote open data. A number of open data portals are in operation in the EU (e.g. publicdata.eu, data.gov.uk, geodata.gov.gr) founded on legislative actions and/or volunteer support, while various EU-funded RTD projects have promoted research on open data (e.g. LOD2).

3.1.4. Critique

As with any new paradigm which gains popularity in the industry and academia, Big Data are not without controversy. Vocal critics of Big Data raise the issues of the underlying theory, fit-for-purpose, and potential abuse.

- **The end of theory**. In a recent article in Financial Times, Tim Harford[12] portrays Big Data as a practice completely ignoring core scientific methodology and tools. As others have also argued[13][14], Big Data supporters believe that insights can be gained from data from *data alone*[15], thus ignoring modeling, assumptions, statistical evaluation. The first ever Big Data use case certainly backs this criticism. Google Flue Trends[16] managed to predict in 2009 the spread of influenza in the USA faster (1 day vs. 5 days) than the Center for Control Disease and Prevention (CDC). Google analyzed user search queries (e.g. search for symptoms, medicine), while CDC collected and analyzed actual data from doctors. This emblematic win was short-lived when in 2013 Google over-estimated the spread by a factor of two[17]. The theory-free and data-driven Big Data analysis is solely based on extracting statistical patterns (*correlation, instead of causation*). While the 'end-of-theory' perspective is not representative for all practitioners, it is important to maintain a clear focus of what Big Data is and how it can help. Proper statistical methodologies, modeling and validating assumptions must be served by Big Data technologies and not be replaced by them.

---

[11] http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/psi_final_version_formatted.docx

[12] http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html?ftcamp=published_links/rss/home_europe/feed//product#axzz2xIRMApw3

[13] http://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory

[14] http://en.wikipedia.org/wiki/Big_data#cite_note-danah2-99

[15] http://www.wired.com/2008/06/the-end-of-theo/

[16] http://www.nature.com/nature/journal/v457/n7232/full/nature07634.html

[17] http://gking.harvard.edu/publications/parable-google-flu%C2%A0traps-big-data-analysis

- **Big data is not for everyone** (*and not everything is Big Data*). As presented in various studies, not all sectors are expected to have the same benefits from Big Data. Further, in most cases the companies reaping the maximum benefits are those that already have massive data available as part of their normal operations (e.g. logs of purchases in an online store, logistics). This data (data *exhaust*[18]) would otherwise be discarded, or produce limited insights. However, for other business sectors Big Data technologies are less suited and (in some cases) useless. The hype over Big Data must be carefully balanced with the actual analysis needs (e.g. size/complexity) and purpose of use. Similarly, Big Data is frequently included as a buzz-word in commercial products and research activities, but without any actual relevance or meaning. The 5Vs mentioned in the previous section should always filter such claims and ground unsubstantiated claims.
- **Potential for abuse** *(Big Brother)*. The capabilities for collecting, integrating, and analyzing massive amounts of data on a world wide scale are considered as a valid threat for user privacy[19]. Indeed, the same technologies used to provide better recommendation services to online customers can be used to profile them and keep track of their activities across their digital activities. Similarly, healthcare data can be used to reduce the mortality rate of cancer patients. The same data and technologies however can be used to exclude consumers from healthcare insurance (e.g. *increased chances of illness due to lifestyle*). For these reasons, privacy protection is an emerging challenge, addressed both on the technical (e.g. anonymization techniques) and policy level (e.g. data retention policies, regulation).

3.2. NoSQL

In this section we present the relational data persistence model, explain its limitations and the new challenges occurring due to new data processing scenarios, introduce NoSQL as a storage model and finally enumerate common NoSQL implementations.

3.2.1. Introduction

Nowadays most of the applications either running on handheld devices or on backend data centers use some kind of persistence storage apart from just writing files on the file system. The prominent mechanism used for storage is relational database management systems (RDBMS).

An RDBMS organizes data in tables of records. Each table (or relation; this term will be used interchangeably) is defined by a set of columns each one representing a distinct attribute of a record. A business entity of any type, such as customer, order, reservation data etc., is stored either in a single relation or may span more than one relation. On top of the storage mechanism, RDBMS offer SQL, a powerful and expressive declarative language for querying and updating data efficiently. Users can implement complex analysis scenarios with SQL by combining data from multiple relations and executing functions on their columns.

For an RDBMS, data integrity is of paramount importance, thus, modern systems support *transactions*. A transaction represents a single logical operation on data that consists of one more independent units of work e.g. transferring funds between two bank accounts, or filling an order and updating warehouse stock. Upon completion, a transaction always leaves the database in a *consistent* state. An RDBMS guarantees that transactions are processed reliably by implementing the *ACID properties*: *atomicity, consistency, isolation and durability*. *Atomicity* requires that if any part of a

[18] http://blogs.wsj.com/cio/2013/11/20/who-owns-your-data-exhaust/
[19] http://blogs.hbr.org/2012/08/dont-build-a-database-of-ruin/

transaction fails, the whole transaction fails and the database is left unchanged. *Consistency* ensures that a transaction leaves database at a valid state, in which all data modifications adhere to all defined rules and constraints. *Isolation* ensures that the concurrent execution of multiple transactions results in the same database state that would have been obtained if all transactions had been executed serially. Finally, *durability* means that once a transaction is completed, its result is permanent, thus, a system failure won't result in any data loss.

Nevertheless, RDBMS have some shortcomings when applied to specific scenarios. An inherent limitation of RDBMS is their inability to provide an intuitive mapping from an application object model to a relational one. This problem is often referred to as object-relation *impedance mismatch*. For instance, representing hierarchical data to tabular relations results in flattening the hierarchy and, thus, lowering querying performance. Moreover, since RDBMS are optimized to manage tabular data, integrating new data types, such as spatial, time series, XML or binary data often leads to inefficient implementations. Many modern RDBMS attempt to remedy this problem by offering separate extensions for each data type which increase the performance. In addition, extensions to the SQL standard are introduced for querying such data. However, such solutions are still tightly coupled with the existing implementations, not allowing for fully optimized solutions.

Also, in an RDBMS the data schema must be known in advance. That is, the table structure, column names and types, must be available before starting to insert data. Evolving the database schema later adds new challenges and implementation changes which may have ripple effects to the whole application architecture. On the other hand, designing a relational schema for being extensible and storing unknown in advance data, often results in implementations that do not make full use of the performance features offered by an RDBMS and thus lead to complex SQL expressions for handling data that may propagate throughout all applications based on such a schema.

Finally, RDBMs where meant to be *vertically scalable*; hence increasing scalability led to powerful but expensive server installations. Moreover, scaling an RDBMS requires some preparation and scheduling. System administrators have to foresee user requirements and scale the system accordingly. Dynamic or unexpected spikes in the number of requests are generally hard to cope with. Nowadays many commercial RDBMs like Oracle and SQL Server can be scaled horizontally (or "scaled out"), but still deployment and configuration remain time consuming. Lately, cloud computing helped to simplify the process even more.

Next, we will introduce *NoSQL databases* and explain how this new technology helps to overcome many of the limitations we have just described.

3.2.2. NoSQL

NoSQL (also named *Not-Only-SQL*) databases attempt to solve the aforementioned problems on storage and processing of Big Data. To achieve that, they are designed with three principles in mind, namely, (a) **simplicity** and schema flexibility, (b) horizontal **scalability** and (c) high **availability**.

Starting with simplicity and schema flexibility, NoSQL databases allow to store different data types, including *unstructured* or *semi-structured* data (e.g. text files, geospatial data, web user click-streams, social media data), and to modify data schema dynamically. Hence development of applications that use such data is greatly simplified. As mentioned in the previous section, an RDBMS would require extensive schema updates in order to accommodate the new data types which may also lead to significant system downtime. In contrast, new data types can be added to a NoSQL database without disrupting existing applications, resulting in increased application development agility. Moving to scalability, a system can be either *scaled up* or *scaled out*. In the former case, scalability is achieved by using more powerful and expensive hardware. In the latter, the same result is reached by using cheaper commodity servers. NoSQL databases are designed to be *distributed* and easily scaled out. Using a set of standard, physical or virtual servers for storing and servicing data, scaling out is as easy as adding new servers to the cluster. Therefore, NoSQL database installations tend to more *cost-effective* than RDBMS implementations which offer similar performance characteristics. Finally, most NoSQL database implementations offer data replication, caching and load balancing out of the box hence allowing for high availability.

NoSQL databases can be classified based on different properties such as performance, scalability, data model etc. The most common classification is based on the underlying supported *data model* resulting in four main categories enumerated below. An extensive description of each category is presented next.

- Key-Value Stores. Data are stored as key value pairs in an associate array. This is the simplest data model available.
- Wide-Column Stores. Store data in records with a dynamic number of columns resembling multi-dimensional key-value stores.
- Document Databases. Store documents of a specific format (e.g. XML, JSON) identified by a unique key. Documents can be queried either by their key or by their contents.
- Graph Stores. Store data that can be represented as graphs where elements have an undetermined number of relations between them.

The difference in the data model allows for some operations to be faster on a NoSQL database than on an RDBMS depending on the processing task under consideration. For example, in a process that requires the use of graph algorithms, a graph NoSQL database may be more appropriate. In addition, most NoSQL implementations *do not fully adhere to the ACID properties*. Relaxing ACID requirements leads to increased performance but removes support for transactions. Moreover, in contrast to using the declarative SQL language, most NoSQL implementations are programmable through *object-oriented* Application Programming Interfaces (APIs) in various languages (e.g. C/C++, Java, C#) or offer RESTful APIs. Hence, the decision of selecting a NoSQL implementation must be taken after careful consideration of the characteristics of the data persistence problem at hand.

Another important aspect of NoSQL databases is the consistency model they employ. In distributed systems, like NoSQL databases, data is partitioned and replicated across many servers in order to achieve availability and fault tolerance. The consistency model dictates how the system reaches a consistent state (i.e. all reads operations get the most current data) after an operation. RDBMS opt for a strict consistency model where every read operation returns the most recently written value. In contrast, most NoSQL databases compromise consistency in favor of high availability and fault tolerance, and implement a relaxed consistency model, named *eventual consistency*. The eventual consistency model guarantees that, if no updates are executed on a particular data item, all replicas will become synchronized and *eventually* all read operations on that item will return the most recently update value.

Before discussing the different types of NoSQL databases, it is important to notice that NoSQL databases can act complementary to RDBMS technology, instead of always trying to replace it. In the past years, developers used to pick a single persistence model and use it throughout the application lifecycle. With the introduction of NoSQL databases, an application may be using more than a single storage model, based on the data usage scenarios and processing requirements. Mixing different storage models will always come with an extra implementation cost, but with tangible cost-efficiency and scalability benefits.

3.2.3. Key-value stores

Key value stores are the simplest form of NoSQL databases. Data is stored in key-value pairs in an associative array with each key being unique. Every implementation defines the format of key values but data values can be anything like text, images, JSON/XML/HTML documents, backups, log files, and more. Different variations exist based on key ordering and type of storage used. For instance, ordered key value stores maintain keys in lexicographic order, hence allowing efficient processing of key range queries. Other implementations favor maintaining data in memory and postpone data persistence on disk when data loss is acceptable by the usage scenarios.

In what follows, a few example implementations of key value stores are enumerated. The list is not complete by any means but illustrates some of the features most implementations offer.

- **Redis**[20]. An advanced key value store that supports multiple key types such as strings, hashes, lists, sets and sorted lists. Redis delivers outstanding performance by implementing an in-memory dataset. Users can opt for disk persistence by periodically dumping memory dataset to disk, or logging all write operations in order to replay them on server start-up. Among other features, atomicity and eventual consistency are supported.
- **Memcached**[21]. A very simple and generic in-memory key value store mainly focused on speeding up web applications by caching data from database calls, external API calls, rendered pages, etc. Memcached supports neither replication nor persistence. Still, performance is maximized and all operations have constant complexity, with every command taking roughly the same time to process every time.
- **Riak**[22]. A NoSQL implementation that offers high availability, scalability and fault tolerance by employing eventual consistency. Value data are versioned and every time the most recent version is returned.

3.2.4. Wide-column stores

Wide column stores store data in records with a dynamic number of columns resembling multi-dimensional key-value stores. Alike to document stores, they are schema free; still their implementation is quite different.

Google's Bigtable [CDG+06] is considered to be the predecessor of most wide column store implementations. Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size of data across thousands of commodity servers. Stored data is indexed by row key, column key and a timestamp that allows multiple versions of the same data. Moreover, data is lexicographically ordered, implementing a persistent multi-dimensional sorted map. The latter allows clients to control the locality of their data through careful choices in their schemas, which accelerates the processing of range queries by decreasing the number of disk accesses and server communication.

A short description of the most popular wide column stores is presented next.

- **Cassandra**[23]. Cassandra is a massively scalable NoSQL database for managing large amounts of structured, semi-structured, and unstructured data and delivering continuous availability, linear scalability, and operational simplicity across many commodity servers with no single point of failure. Data is distributed transparently across all nodes that participate in a database cluster (or "ring" as called in Cassandra

[20] http://redis.io/

[21] http://www.memcached.org/

[22] http://basho.com/riak/

[23] http://cassandra.apache.org/

terminology). Moreover, if required, Cassandra provides built-in and customizable replication, which stores redundant copies of data across nodes that participate in a Cassandra ring.

- **HBase**[24]**.** HBase is a distributed, scalable, versioned, non-relational data store able to host very large tables sizing to billions of rows with millions of columns, hosted atop a cluster of commodity hardware. HBase is modelled after Google's Bigtable. Just as Bigtable leverages the distributed data storage provided by the Google File System, HBase provides Bigtable-like capabilities on top of the Hadoop Distributed File System (HDFS), a distributed file system designed to run on commodity hardware with linear and modular scalability.
- **Apache Accumulo**[25]. Accumulo extends Google's Bigtable design and is built on top of several Apache projects, namely, (a) Hadoop, a framework that allows implementation of distributed processing of large data sets across clusters of computers using simple programming models, (b) Zookeeper, a service that enables highly reliable distributed coordination, and (c) Thrift, a framework for scalable cross-language services development. Accumulo features a few novel improvements including cell-based access control and a server-side programming mechanism that can modify key-value pairs at various points in the data management process.

3.2.5. Document databases

The main concept in document databases is the notion of "document", also known as semi-structured data. Data is encapsulated inside a document, usually encoded in a well-known text format like XML, JSON, YMAL, etc. as well as in binary formats like Portable Document Format (PDF), Microsoft Office documents (e.g. Word, Excel). Each document is identified by a unique key like in key-value stores. Still, a distinguishing feature of document databases is that not only keys can be indexed and queried, but also document contents.

Document databases can be compared to relational databases with collections of keys being analogous to relations and documents to records. However, the main difference is that for each relation there is a rigid predefined schema to which every record of the relation must adhere, while a document can have an arbitrary number of fields. For instance, the following two example documents are both valid for a document database which stores data in XML format. Notice that each "record" (document) is neither required to share a specific set of elements nor to contain empty elements for fields whose value is not available.

```xml
<book>
   <title>Transaction Processing: Concepts and Techniques</title>
   <authors>
      <author>Jim Gray</author>
      <author>Andreas Reuter</author>
   </authors>
   <language>English</language>
</book>
```

```xml
<book>
   <title>Introduction to Algorithms</title>
   <authors>
      <author>Thomas H. Cormen</author>
      <author>Charles E. Leiserson</author>
      <author>Ronald L. Rivest</author>
      <author>Clifford Stein</author>
   </authors>
   <price>79.13</price>
   <publisher>The MIT Press</publisher>
</book>
```

[24] http://hbase.apache.org/

[25] http://accumulo.apache.org/

Different implementations of document databases may offer specialized query languages for querying document content (like XQuery) apart from programmatic APIs. Depending of the persistence and indexing scheme a document database implements, significant performance variations should be expected.

At the time of this writing, two of the most popular document databases are MongoDB and CouchDB.

- **MongoDB**[26]. MongoDB is currently the most popular document database. Data is stored in the form of binary JSON-like documents with dynamic schemas. More precisely, MongoDB documents are BSON documents. BSON is a binary encoded JSON with additional data type support, e.g. date and timestamp types are supported. Atomicity is supported at document level and only eventual consistency is implemented.
- **CouchDB**[27]. Similar to MongoDB, CouchDB stores data as JSON documents. Moreover, it offers limited ACID functionality through multi-version concurrency control which allows all readers to see a consistent snapshot of the database during read operations without being locked out by writes.

3.2.6. Graph stores

A graph store is a database for storing data that can be represented as elements (nodes) with an undetermined number of relations (edges). Such data can represent road maps, network topologies, traffic data, social network interactions, web page navigations etc.

A graph store provides *index-free adjacency*, which means that every element in the database contains direct links to its adjacent elements. Hence, every element can access all the nodes to which it is connected without the need of an index lookup. This feature accelerates the execution of graph theory algorithms on the stored data and allows graph databases to query associative data sets faster than existing RDBMS. In contrast to RDBMS, no expensive join operations are required for accessing relation information which usually is one of the most important information types.

One of the most popular graph database implementation is **Neo4j**[28]. Neo4j is a scalable, native graph database with full ACID support. It provides a query language, namely Cypher, for querying graphs as well as programmatic bindings for most popular programming languages such as Java, .NET, JavaScript, Python, etc. A more specialized version of graph databases are the RDF stores (e.g. Virtuoso[29], Sesame[30]), which are useful for storing and querying RDF triples. In RDF stores the predicate can be interpreted as a relation (edge) between two nodes representing the subject and object. Nevertheless, RDF stores offer more advanced functionality such as support of SPARQL, a SQL-like query language for RDF, reasoning and schema flexibility.

3.3. Map-Reduce

Processing or generating big datasets is usually a computationally expensive job that takes a long time to complete. Most of the time, such jobs are executed using *parallelism* and *distributed computing* in order to achieve timely

[26] http://www.mongodb.org/
[27] http://couchdb.apache.org/
[28] http://www.neo4j.org/
[29] http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/
[30] http://www.openrdf.org/

responses. Implementing such solutions requires significant effort for splitting the job in manageable tasks that can be processed efficiently by a single server, communicating data between servers, monitoring remote processes, handling software errors or hardware failures, composing the final result from partial responses, etc. It is not unusual that most effort is spent in implementing parallelism and coordinating the execution process than implementing the algorithm that solves the problem under consideration.

MapReduce [DG04] is a programming paradigm and implementation framework that accelerates the development of such solutions by abstracting the execution process management, thus allowing developers to concentrate on implementing an algorithm for their problem. Therefore, programmers can develop efficient, highly distributed solutions without prior knowledge of parallel and distributed systems.

The programming paradigm of MapReduce requires that *at least two methods* are provided for solving a problem, namely, a **Map** method that processes a key-value pair and generates a set of intermediate key-value pairs (usually in a different data domain), and a **Reduce** method that is applied on all intermediate pairs with the same key and returns zero, one or more values.

The associated implementation framework requires the specification of some initialization parameters and thereafter takes care of all execution details including partitioning input data, scheduling tasks to a set of servers, performing load balancing, monitoring process execution, and composing the final result. Moreover, in order to increase parallelism and fault tolerance, a distributed storage system is used and data locality is exploited during data partitioning and scheduling.

A simple word-count MapReduce program is presented below using pseudo code. The program parses a set of text documents stored on several servers and computes the number of occurrences of each word:

```
// key: name of a particular file
// doc: contents of this file
map(string key, document doc) {
       // iterate over all words in document
       for each word w in doc {
              return (w, 1);
       }
}
// key: unique word
// values: iterator that enumerates all intermediate key value pairs (key, 1)
reduce(string key, iterator values) {
       int result = 0;
       // Get next value from the iterator until the list of values is exhausted
       while( values.next()) {
              // Aggregate all values. The value of each key value pair generated by
// the map method is equal to 1
              result += ParseInteger(values.current());
       }
       return result;
}
```

The map function accepts key-value pairs that represent filenames and their corresponding contents and returns a key-value pair for each word it parses with the default value of 1. After the framework orders the intermediate keys, it sums all occurrences for each unique word by calling the Reduce method.

The simple example above describes the work required by a developer. Next, we examine the implementation details of the MapReduce framework.

During the execution of a MapReduce program, the data input is partitioned in M sets. Each set is processed in parallel by different servers, generating intermediate key-value pairs by calling the Map method on each element of the input set. After computation of the Map method, the intermediate results are partitioned into R sets by partitioning the intermediate key using some partitioning function (e.g. hash (k2) modulo R). Finally, the R sets of intermediate results are processed in parallel where the Reduce method is applied. Both parameters M and R are specified by the user. The system has to execute M Map tasks and R Reduce tasks resulting in O (M+R) scheduling decisions overall.

The execution steps of a MapReduce program can be summarized as follows:

- The system splits the input data into M sets. The size of each set can be controlled by the developer during the program initialization. Typically the set size varies from 16 to 64 megabytes. Selecting a small set size results in manageable input size for a single server and at the same time increases parallelism by exploiting locality of the underlying distributed storage system.
- The framework initializes multiple copies of the program on a set of servers. Those servers are referred as *workers*. One of them is responsible for scheduling the Map and Reduce tasks and is called the *master*.
- Workers that are assigned Map tasks read a set of input data and apply the Map method to each key-value pair in it. The intermediate results are buffered in memory and periodically stored locally after they have been partitioned in R files (or sections) using the partitioning function. Runtime information about the partition location and execution progress is transmitted to the master. In total, M tasks must be processed and M*R sections are generated.
- Workers that are assigned Reduce tasks are informed by the master about the location of intermediate results. After all M_i intermediate results for a specific index in range [1, R] are fetched, they are sorted by the intermediate key and the Reduce function is called for each unique key and the corresponding set of intermediate values associated with it. The result is stored to a final output file in the distributed storage system. At the end of reduce execution, R results files are generated.

Another important aspect of the MapReduce framework is its *automatic fault tolerance*. On the process described above, if a Map worker fails, the task assigned to it is rescheduled, even if it has completed successfully, since the intermediate results are stored locally. On the contrary, a Reduce worker is rescheduled only if not already completed, since the final result is always written on the distributed storage system. In the case the master worker fails, the program aborts resulting in a single point of failure architecture. An alternative solution is to have the master write periodic checkpoints with all the execution status information about scheduling and intermediate result locations. In case of a failure, a new master is restarted at the latest checkpoint.

Finally, additional extensibility points are offered, allowing developers to customize the framework behaviour at runtime. Such features include:

- **Input**: Implement custom readers for reading input data e.g. accessing records in a relational database or key-value pairs from a NoSQL database.
- **Partition**: In general the default partitioning function generates well balanced partitions. However, there are scenarios that users may require to group the results differently. For example, if the output results are key-value pairs where the key is a URL, it may be preferred that URLs from the same domain are grouped in the same output file. For this case, a hashing function like hash(domain(URL)) may be preferable
- **Combine**: As shown in the pseudo code above, sometimes the intermediate results may contain many values with the same intermediate key. For instance, the pairs ("the", 1), ("a", 1), ("an", 1) may occur thousands of times in comparison to less frequently used words. Transferring these values from the local disks of the Map workers to a Reduce worker and sorting them before running the Reduce method, causes both additional network traffic and computational cost. To remedy this problem, a Combine function can be provided that does partial merging of the data locally at the Map worker before it is sent over the network.
- **Output**: Similar to the custom readers, developers can provide custom writers for writing output files to different formats in order to be consumed by other applications.

3.3.1. Hadoop

Apache Hadoop[31] is a framework for distributed processing of large datasets across a cluster of servers using a simple programming model. Instead of relying on expensive hardware for achieving high availability and scalability, the framework handles hardware and software errors at the application level, thus delivering highly available and scalable computing over a cluster of commodity servers. Hadoop consists of four core modules:

- a set of common utilities that support the other Hadoop modules,
- a distributed file system (HDFS),
- YARN [VM+13], a framework for task scheduling and resource management, and
- a MapReduce implementation based on YARN for parallel and distributed processing of large datasets.

Next we present HDFS, discuss the Hadoop MapReduce implementation and enumerate the main components of YARN.

When data files are loaded in a Hadoop cluster, the Hadoop Distributed File System (HDFS) splits them in smaller blocks of a fixed size and saves multiple copies of each block across several servers in order to guarantee data availability in case of hardware failures. The servers which store each block are chosen randomly on a block-by-block basis. These individual servers are referred to as **DataNodes**. Even though file blocks are replicated and distributed across several servers, the process of partitioning the files and distributing the blocks is transparent to the end users. In order to achieve that, HDFS stores metadata for each file and all of its blocks. In addition, a monitoring service replicates data as needed when system failures occur in order to sustain the system scalability. Hence, if a small percent of servers, e.g. a server rack, goes offline the system performance is affected only by a proportional factor.

[31] http://hadoop.apache.org/

The default block size in HDFS is 64 megabytes. Since HDFS expects to store very large files, a large block size decreases the overall size of metadata stored. Moreover since MapReduce programs usually scan large input files, keeping large amounts of sequentially stored data on a disk results in fast reads. Alternatively, increasing the block size results in less parallelization of Map tasks in a MapReduce job.

The metadata for file partitioning and replication is handled by a single server, namely, the **NameNode**. When a client requests a file, it first contacts the NameNode to access the block locations for the specific file and then retrieves data from the corresponding DataNodes. As metadata size is relatively low (only information about filenames, block locations per file and permissions are kept), the NameNode stores this information in the main memory and can serve a high number of clients with a minimum overhead.

Since metadata is stored on a single server, it is important that it is stored reliably. A NameNode failure is more severe than that of a DataNode since it renders the whole file system inaccessible. In order to scale naming service horizontally, it is possible to declare multiple NameNodes with their corresponding namespaces (file systems). Each NameNode has access to all available DataNodes. NameNodes are independent and don't require coordination with each other. Each NameNode is a single point of failure for the associated file system only. Therefore, the overall availability of the system is increased.

HDFS is used for implementing MapReduce in Hadoop, in the same manner BigTable [CDG+06] is used for implementing Google's MapReduce. The implementation is fairly similar and most of the features described in the MapReduce section apply to the Hadoop MapReduce implementation. The Hadoop framework schedules map and reduce tasks using data locality knowledge from HDFS. Thus computation in moved to the data, instead of moving the data to the computation which results in high performance and less network traffic.

Moreover, Hadoop provides a mechanism for managing batches of jobs with dependencies among them. Instead of submitting a single job, users can declare multiple jobs with dependencies and submit them as a single unit of work. A job won't start before all its dependencies have completed successfully.

Another important aspect of Hadoop MapReduce is that job scheduling and resource management is separated from the MapReduce programming model implementation and handled by another framework, namely, YARN (Yet Another Resource Negotiator). YARN is agnostic to the applications that request resources and allows Hadoop to be used for implementing other computational models apart from MapReduce such as graph processing, machine learning etc.

YARN uses a single **ResourceManager** for managing collectively the resources of the whole cluster and one **NodeManager** per server for managing the server's local resources. The resources on each server are organized in leases (containers) that represent CPU, memory, disk space, network bandwidth etc. The ResourceManager has two main components, namely, the **Scheduler** and the **ApplicationsManager**.

The ApplicationsManager is responsible for accepting job requests, initializing job execution and monitoring their status in case a restart is required when a failure occurs. The Scheduler is responsible for allocating resource containers to running applications (jobs) and enforcing constraints on the resources utilization. The policy for sharing resources among various applications is pluggable and extensible. An example of such a pluggable scheduler is the FairScheduler that assigns resources to applications such that all applications get, on average, an equal share of resources over time.

The ResourceManager is ignorant of the semantics of each resource allocation requested by an application. For each application (job) running on the cluster an ApplicationMaster is assigned. **ApplicationMaster** is actually a framework specific library that implements a particular programming model e.g. MapReduce and is responsible of negotiating appropriate resource containers from the Scheduler, generating a physical execution plan, monitoring and tracking execution progress and handling execution errors. Before starting a new job, a container must be acquired from the ResourceManager for launching the ApplicationMaster itself. Afterwards, the ApplicationMaster negotiates resources from the ResourceManager and once sufficient resources are acquired, it works with the one or more instances of NodeManager for executing its jobs.

Besides the main components of Hadoop, a few related projects are in development:

- **Ambari**[32] is a web based user interface tool that simplifies the provisioning, management and monitoring of Hadoop clusters. It allows system administrators to install and configure Hadoop services across multiple servers in a cluster, manage installed services e.g. start, stop, configure etc., monitor cluster status and send notifications in case administrator's attention is required e.g. remaining disk is low.
- **ZooKeeper**[33] is a centralized service for storing configuration, naming and group information and providing distributed synchronization. Such functionality is usually required when developing distributed applications. Using ZooKeeper simplifies the development of such applications and makes the whole process less error prone. The service itself is highly distributed and reliable.
- **Tez**[34] is an application framework based on YARN that generalizes the MapReduce paradigm to a more powerful framework for executing a complex directed acyclic graph of tasks. Tez eliminates unnecessary tasks and minimizes access to HDFS in order to speed up data processing.

3.3.2. Stratosphere

Stratosphere (now Apache Flink[35]) is a data processing system for analyzing large datasets. Instead of building its functionality on top of an existing MapReduce framework, Stratosphere implements its own job execution runtime. Therefore, it can be used either as an *alternative* to Hadoop MapReduce component or as a *standalone processing system*. When used with Hadoop, Stratosphere can access data stored in HDFS and request cluster resources from the YARN resource manager.

Stratosphere extends the MapReduce programming model with additional operators, also called transformations. An operator consists of two components, a user-defined function (UDF) and a parallel operator function. The operator function parallelizes the execution of the user-defined function and applies the UDF on its input data. The data model used by Stratosphere operators is record based instead of the key-value pair model used by MapReduce. Still, key-value pairs can be mapped to records. A collection of records is referred to as dataset. All operators will start working in memory and gracefully fallback to external memory algorithms once memory resources become low. The new operators represent many common data analysis tasks more naturally and efficiently

A short description of the available operators as presented in the Stratosphere documentation follows.

[32] http://ambari.apache.org/
[33] http://zookeeper.apache.org/
[34] http://tez.incubator.apache.org/
[35] http://flink.incubator.apache.org/

- **Map**: The Map transformation applies a user-defined function on each element of the input dataset. It implements a one-to-one mapping, that is, exactly one element must be returned by the function. This behavior differs from that of the classic Map operator.
- **FlatMap**: The FlatMap transformation applies a user-defined function on each element of the input dataset. This variant of a map function can return arbitrary many result elements (including none) for each input element. The behavior of this operator matches that of the classic Map operator.
- **Filter**: The Filter transformation applies a user-defined function on each element of the input dataset and retains only those elements for which the function returns true.
- **Project**: The Project operator allows the modification of the fields of a tuple. A tuple is an ordered list of fields. Users can shuffle, add or remove fields. Project operator does not need the definition of a user-defined function.
- **Reduce** on grouped dataset: A Reduce operator that, when applied on a grouped dataset, it reduces each group to a single element. For each group of input elements, the user-defined function successively combines pairs of elements into one element until only a single element for each group remains. There are different variations for this operator. For example, users may define an additional function for extracting the key used for grouping from each element.
- **GroupReduce** on grouped dataset: A transformation that, when applied on a grouped dataset, it calls a user-defined function for each group. The difference between GroupReduce and Reduce is that the user defined function gets the whole group at once instead of a pair of elements at a time. The function is invoked with an iterator over all elements of a group and can return an arbitrary number of result elements. This operator resembles the classic Reduce operator.
- **Reduce** on full dataset: Applies a user-defined function to all elements of the dataset. Pairs of elements are subsequently combined into one element until only a single element remains.
- **GroupReduce** on full dataset: Applies a user defined function to a dataset by iterating over all the elements of the dataset. An arbitrary number of result elements is returned.
- **Aggregate** on grouped tuple dataset: Supports min, max and sum aggregation operations. The aggregate transformation can only be applied on a dataset of tuples.
- **Join**: Joins two datasets into one dataset. The elements of both datasets are joined on one or more keys which can be specified either by a user defined function or by field indexes, if elements are tuples.
- **Cross**: The Cross transformation combines two datasets into one dataset by building a Cartesian product. The Cross transformation either calls a user defined function on each pair of elements or applies a projection
- **CoGroup**: The CoGroup transformation jointly processes groups of two datasets. Both datasets are grouped on a defined key and groups of both datasets that share the same key are passed together to a user-defined function which iterates over the elements of both groups. If for a specific key of any dataset there are no matching elements from the other dataset, an empty group is passed to the user function.
- **Union**: Produces the union of two datasets, which have to be of the same type.

Stratosphere allows to model job processing as *directed acyclic graphs (DAGs)* of operations, which is a more flexible model than MapReduce, in which Map operations are strictly followed by Reduce operations. The combination of various operations allows for data pipelining and in-memory data transfer optimizations, which increase performance by drastically reducing disk access and network traffic. Moreover, Stratosphere supports highly efficient iterative algorithms, which are very important for Data Mining, Machine Learning and Graph processing, since such jobs often require looping over the working data multiple times. Implementing such jobs with MapReduce is quite expensive since data are transferred between iterations by using the distributed storage. In contract, Stratosphere supports iterative algorithms in its core.

Stratosphere offers powerful APIs in Java and Scala. The Stratosphere optimizer compiles the user programs into efficient, parallel data flows which are executed on a cluster or a local server. The optimizer is independent of the actual programming interface and supports cost-based optimization for selecting operator algorithms and data transfer strategies, in-memory pipelining of operators, data storage access reduction and sampling for determining cardinalities.

3.3.3. Other Map-Reduce frameworks

In this section we present short descriptions for various systems that implement the MapReduce programming model.

- **Apache Hive**[36] is a data warehouse infrastructure built on top of Hadoop. It allows for querying data stored on HDFS via HiveQL, an SQL-like language that gets translated to a set of MapReduce jobs. HiveQL language also allows programmers to plug in existing mapper and reducer implementations when it is inconvenient or inefficient to express the required logic in HiveQL. Despite providing SQL functionality, Hive does not provide interactive querying. Instead it only runs batch processes on Hadoop and is best suitable for batch jobs over large sets of append-only data (like web logs). Integrating Tez with Hive will allow users to execute queries more efficiently. Tez facilitates simpler, more efficient query plans, which translates to significant performance improvements due to the pipelining of reduce tasks.
- **Apache Pig**[37] supports large dataset analysis by offering a high level language for expressing data analysis programs and an infrastructure for evaluating these programs. The main property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large datasets. Currently, the infrastructure consists of a compiler that produces sequences of MapReduce programs that can be executed in Hadoop. Just like Hive, integrating Pig with Tez will increase efficiency.
- **Apache Spark**[38] is a fast and general purpose cluster computing system that provides similar functionality to Stratosphere. It has an advanced DAG execution engine that supports cyclic data flow and in-memory computing. It provides APIs in Java, Scala and Python and offers more than 80 high level operators that make it easy to build parallel applications. Spark can also be used interactively from the Scala and Python shells. On top of Spark processing framework, a stack of high levels tools is provided including Spark SQL for querying structured data: MLlib, a library for machine learning that takes advantage of Spark's iterative computation efficiency; GraphX for graph processing; Spark Streaming for developing streaming applications; and Shark that offers increased performance and advanced analytics to existing Hive deployments. All these frameworks can be combined seamlessly in the same application. Sparks executes programs up to 100x faster than Hadoop MapReduce in memory, or 10x faster on disk.

[36] http://hive.apache.org/
[37] http://pig.apache.org/
[38] http://spark.apache.org/

3.3.4. Cloud data management

Hadoop's power comes from its ability to perform work on a large number of machines simultaneously. Using a small cluster consisting of a few servers only, will not yield significant performance improvements. At the same time, deploying and maintaining a larger cluster is challenging for most small organizations or companies by means of IT costs. Moreover, most developers experiment and develop applications with Hadoop locally due to the absence of a larger cluster. Development is not a problem since operations on a cluster consisting of a few nodes are functionally equivalent to those on a cluster of 100 or more nodes. Nevertheless, benchmarking and fine-tuning the implemented solution requires the presence of an appropriate cluster setup. Cloud computing helps both companies and developers to deploy Hadoop on low cost clusters by alleviating the burden of installing and maintaining a large cluster on local premises.

Next we present two cloud services, namely, Amazon Elastic Compute Cloud and Microsoft Azure, which offer flexible deployment and billing schemes.

3.3.4.1. Amazon Elastic Compute Cloud

Amazon Elastic Compute Cloud (Amazon EC2[39]) is a web service that provides resizable processing capacity in the cloud through a simple interface that allows developers to easily obtain, configure and control processing capacity with minimal effort. As computing requirements change, Amazon EC2 supports simultaneous instantiation of multiple servers within minutes, allowing it to scale fairly fast. At the same time, users are billed only for the capacity they are actually using. The billing plans are very flexible and users can be billed even hourly based on the computing capacity used. Thus, implementing applications on Amazon EC2 offers both scalability and cost efficiency.

When creating a new server (or instance), Amazon EC2 allows developers to select the operating system, the memory configuration, assigned CPU resources, available storage and software packages to be installed. Furthermore, since everything is controlled and configured using web service APIs, developers can even build applications that automatically scale up or down depending on the computing load.

Amazon EC2 offers high reliability by allowing instances to be replaced swiftly. Moreover, secure and robust network functionality is provided. All instances are located inside a virtual private cloud and only the instances that are specified by the administrators are made public.

Finally, applications that are based on Amazon EC2 can access other Amazon web services like:

- Amazon Simple Storage Service (Amazon S3): Provides a web service interface for storing and retrieving data over the web.
- Amazon Relational Database Service (Amazon RDS): Provides a web service for operating a relational database in the cloud.

  Amazon SimpleDB: A highly available, flexible, scalable and easy to administrate NoSQL database served as a web service that enables users to focus on the application development instead of database administration. SimpleDB supports flexible data schemas that can be evolved dynamically alongside with the user applications. Moreover, users can choose between strict and eventual consistency depending on

[39] http://aws.amazon.com/ec2/

their application needs, thus, balancing performance and consistency. The "set and forget" nature of SimpleDB makes it a perfect candidate for scenarios like storing application logs, capturing environmental measurements (e.g. temperature, humidity, etc) and tracking geolocation information.

- Amazon Simple Queue Service (Amazon SQS): A reliable and scalable message queuing web service for decoupling cloud application components.
- Amazon EMR: A web service based on Hadoop that makes it easy to quickly and cost-effectively analyze vast amounts of data. With Amazon EMR, a Hadoop cluster can be configured in minutes in contrast to the hassle required to configure the same cluster manually. After configuration, processing capacity can be altered based on the system load and processing requirements. Moreover, EMR supports Hadoop tools like Hive, Pig and HBase for data analysis. Amazon EMR can be used for a variety of applications that require batch analysis of vast datasets, including log analysis, web indexing, data warehousing, machine learning, financial analysis, scientific simulation and bioinformatics.

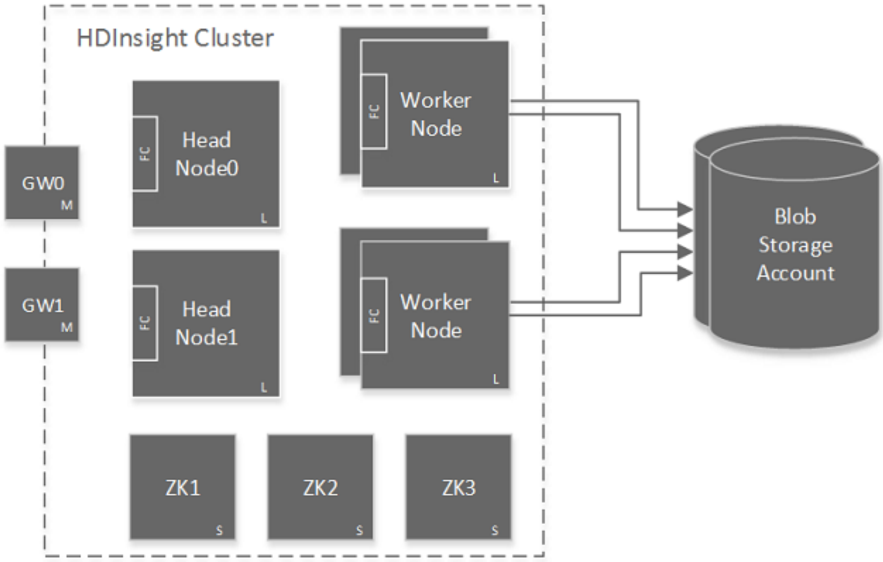
3.3.4.2. Microsoft Azure

Microsoft Azure[40] is a flexible cloud platform that enables users to quickly develop, deploy and manage highly available applications using the programming language and framework of their preference without focusing on the infrastructure. Likewise to Amazon EC2, Azure can scale applications to any size by allowing the provisioning of resources within minutes. Available resources can elastically grow or shrink based on application load. Moreover, users pay only for the resources used by their applications.

Azure features and services are exposed using REST protocols. Users can store data using relational SQL databases, NoSQL table stores and unstructured blob stores. Optionally, Hadoop and business intelligence services can be used to data-mine stored data. In addition, a reliable and robust messaging service enables the development of scalable distributed applications.

Finally, Azure offers HDInsight[41], an Apache Hadoop-based service in the cloud for data analytics. HDInsight allows to build a Hadoop cluster in minutes, perform data analysis, analyze unstructured data in Microsoft Excel and mash up results from HDInsight with data from internal, external, relational and non-relational sources. More specifically, Azure HDInsight makes the Hadoop HDFS and MapReduce framework, along with other Hadoop related projects like Hive, Pig and HBase, available in a simple, scalable and cost effective environment. First of all, HDInsight increases availability by instantiating two NameNodes (head nodes) instead of a single one, thus, removing the single point of failure. ZooKeeper nodes (ZK) are also created to ensure that DataNodes (worker nodes) and gateway nodes have knowledge of the active NameNode as shown in the diagram below.

[40] http://azure.microsoft.com/en-us/
[41] http://azure.microsoft.com/en-us/solutions/big-data/



Picture 41: Azure HDInsight high availability http://azure.microsoft.com/en-us/documentation/articles/hdinsight-high-availability/

Another important feature of HDInsight is how data is stored and managed. Instead of using HDFS only, HDInsight also uses Azure Blob Storage (ABS), a robust, low cost, long-term and general purpose storage solution, as the default file system. ABS is optimized for data storage while HDFS is optimized for data processing. Using ABS as the default file system allows users to delete Hadoop nodes when computation jobs are completed without losing data. Additionally, ABS provides a full-featured HDFS file system interface, thus, enabling Hadoop components to operate directly on the data that it manages. The additional performance cost that occurs for not having compute and storage located on the same server is mitigated by the way the Hadoop clusters are provisioned i.e. servers close to the stored data are preferred. Moreover, using ABS offers additional benefits:

- More applications gain access to the data through the Blob Storage REST APIs.
- Data can be archived in the blob storage and the compute clusters can be safely deleted when they are not needed anymore.
- Storage cost is reduced since compute nodes are more expensive than storage nodes.
- Blob Storage scales out automatically. In contrast, for HDFS to scale out, additional nodes must be created.

All these features make HDInsight an attractive solution for big data scenarios like monitoring supply chains in retail, suspicious trading patterns in finance, demand patterns for public utilities and services, air and water quality from arrays of environmental sensors, or crime patterns in metropolitan areas. These kind of activities require ad hoc analysis, in batch mode, of unstructured datasets, which are not updated frequently. HDInsight Hadoop-based implementation handles such scenarios gracefully.

3.4. Data Analysis

A main goal of DAIAD is the efficient and scalable analysis of vast amounts of water consumption measurements, in the form of time-series, which comprises the following tasks: (a) Recognizing consumption patterns in several granularities (e.g. users, groups of similar users, neighborhoods, cities) (b) Correlating these patterns with several factors that might influence water consumption (e.g. seasonality, demographics, weather); (c) Producing personalized and aggregate recommendations for behavioral changes (e.g. consumption reduction) to users or groups of users; (d) Providing sophisticated and problem-specific algorithms for consumption prediction and modeling in large scale (*city-level*) to be utilized by groups of consumers for collaborative and personalized consumption adjustment, as well as by water demand experts for modeling and policy adjustment.

In what follows, we present state of the art works touching each of the above topics. We begin by presenting two paradigm studies, on energy and water consumption respectively. Although much more confined and small-scaled, compared to our vision on the project, these studies provide some firm guidelines and intuitions about decisions and attributes, regarding the development of large scale consumption analysis frameworks. Then, we present a series of works on pattern recognition and event/anomaly detection on time series, to demonstrate the plethora of machine learning algorithms and models that can by adopted, enhanced and extended to fit our problem's requirements. Also, we provide a brief description of several recommendation and personalization approaches and explain the specificities of our setting, with respect to these techniques. Finally, we discuss anonymization works that can guide the development of safe infrastructures regarding the privacy of consumers.

3.4.1. Consumption Analysis Paradigms

Next, we describe two consumption analysis paradigms performed on energy and water consumption respectively. Apart from enumerating the major points and findings of these studies, we further discuss and extend them in order to identify some important challenges to be faced during DAIAD.

3.4.1.1. Energy Consumption Analysis

The white paper of [SW13] presents a study, performed by KNIME.com AG on analyzing smart meter, energy consumption data in order to identify similar consumption behaviors and predict future consumptions. It is a small scale study, involving 6,000 households and providing consumption measurements of maximum granularity of 30 minutes, on the total consumption of each household, for one year. It was performed in Ireland during 2009 and 2010 by the Irish Commission for Energy Regulation (CER) and with the involvement of several energy stakeholders. According to the authors, the aim of the study is: (a) To provide the ability of customizing energy contract offers for customers-consumers (via recognition of consumption behavior patterns) and (b) To predict future consumption patterns to assist electricity companies in policy making (via consumption prediction models). The study works also as a proof of concept for the company's analytics platform; it adopts a simple analysis workflow, as well as two generic pattern recognition and prediction models to produce its results. However, it provides some problem-specific guidelines and intuitions on the analysis workflow and the effectiveness of the applied models that could be useful in DAIAD's framework. Next, we identify and shortly discuss the major points and outcomes of the study.

Granularity. The first premise presented in the study revolves around the granularity of the data (w.r.t. both users-consumers and time) and how this factor affects the effectiveness of the methods and the accuracy of the results. On user granularity, the major claim is that focusing on a medium granularity level, namely forming and analyzing groups of similar users is the most advantageous strategy: Analyzing the total energy usage would produce too general results to be reliable, while analyzing the consumption of each user-household separately would be too computationally-costly and would yield too overspecialized findings. On time granularity, it is claimed, as well as indicated through experiments, that seasonality of consumption can be identified in several levels: day hours, week days, year seasons. For example, for specific (family) households, the consumption during the night is distinctively lower than during the day. On the other hand, student households or small businesses can present the opposite behavior: high consumption during the night. Another example is the significant drop in consumption in Sundays or during Christmas breaks.

Combination of models. The study examines two models for the two distinct tasks it assesses: A *k-means* clustering algorithm for identifying similar users w.r.t. their consumption behavior and a *regression* model for predicting future consumptions. While the sequential application of these two techniques yields accurate results, the authors point out the need for combining several models in order to refine the outcomes. Indeed, the specific problem (and, in general, analyzing time series of consumption measurements irrespectively of the nature of the measurements) involves considering several context parameters: seasonality, weather (temperature), demographics, geographic location, etc. Taking into account all these parameters within the analysis process requires combining, fine-tuning and specializing several machine learning and data mining models and, possibly, selecting different models according to the specificities of each task or even the context of each use case scenario. For example, one could apply a standard regression model to predict the consumption of a user for the next week, given the user's consumption history. However, if a collaborative model has been trained, and similar users have been identified, then (a) the consumption prediction of the model for the user could be refined and (b) the user could be recommended consumption behaviors that have been recently adopted by similar users in order to improve her consumption behavior. On top of that, the applied models should be chosen and parameterized according to model-theoretic guidelines possibly available by water demand management experts, increasing, thus, their expected accuracy.

Feature selection. Time series data on user consumption pose several challenges on applying state of the art machine learning algorithms for pattern recognition, classification, regression, etc. Apart from the actual measurement values, one must take into account their correlation with previous values, their fluctuations, seasonality, effect of context factors, etc. Thus proper features that represent and correlate different aspects of the data need to be defined and explored. This is also identified in the study, where proper features are defined to be used in the process of recognizing similar consumption behavior patterns between users. These features include average and percentage values of consumed energy during five different day segments, average yearly, monthly, weekly, daily, and hourly consumed energy, etc.

Data volume and processing choices. The described analysis consisted of several processing steps: Data extraction and loading; Data transformation; Data summarization and visualization; Data mining and machine learning; Data analytics. Some of them might be relatively lightweight, while others might be very resource-demanding, requiring large processing power and/or distributed processing. The general point is that, in such complex processing workflows, where several processing and analysis tasks must run, it is often the case that there is not one-size-fit-all solution: Some very demanding tasks, e.g. mining terrabytes of data to discover a pattern, might require a distributed solution where tenths of machines combine their processing power under a common framework (Map-Reduce). On other cases, however, e.g. when we want to classify a consumption pattern according to an already trained classification model, it would be much faster to utilize a single machine. Thus, carefully selecting the processing models and the algorithms/tasks to run on them is crucial for optimizing the efficiency and scalability of such frameworks.

3.4.1.2. Water Consumption Analysis

The second study [NL+11] focused on analyzing water consumption patterns and providing analytics and insights in several granularity levels (per users/groups/experts, historical, aggregated data, consumption forecasting, etc.). Although it was performed in a smaller scale (~300 volunteered households and 15 weeks duration), it conducted a more thorough examination of parameters influencing consumption, user behaviors and problem modeling. It was orchestrated by IBM and the city of Dubuque, US and it was based on smart meter measurements on household water consumption. The main goals of the study were: (a) To establish a new baseline for water consumption (using smart water meters), educate citizens about water conservation and reduce overall water usage; (b) To test the hypothesis that informed and incented citizens would be able to conserve water more efficiently; (c) To serve as a template for sustainable communities that desire to engage their constituents in resource conservation whether it is water, energy or any other natural or produced resource. A major factor in the performed experiment was a portal where users could overview several statistics on their consumptions, socialize, compare, provide/get feedback with other, similar users and experts could review aggregate consumption data and analysis results.

Study roadmap. The 15-week water pilot study was conducted in two phases with two groups of about equal size.

- Phase 1 lasted eight weeks. During it, a Pilot group of 151 users was given access to the portal, to printed weekly reports, and to support personnel. They could receive leak alerts, monitor/analyze their own water usage, patterns and trends, compare their usage to that of others, and collaborate online via chat and weekly team-based contests. During this period a Control group of 152 users had no access to any of the above mentioned. The purpose of Phase 1 was to allow a direct contrast between the amount of water reduction by the Pilot and Control groups.
- After a one-week break, Phase 2 commenced and lasted for six weeks. During Phase 2, the Control group of 152 users was also given access to the same functionality as the Pilot group – i.e., the portal, printed weekly reports, and support personnel, and so on. The purpose of Phase 2 was to get further information on the use of the Water Portal, with the 152 Control group volunteers using the Portal along with the 151 members of the Pilot group.

Specifically, during the study, the users were able to review their consumption in several granularity levels (hours, days, weeks), get alerted about unusual consumption patterns (e.g. leaks), compare their consumption with other (similar) households, share information (e.g. tips, best practices), participate in social games/contests with the aim of consumption reduction. In parallel, the experts, that had the ability to view aggregate (anonymized) consumption data from all users, could monitor aggregate consumption statistics, identify spatial and temporal anomalies in consumption, forecast future consumptions and assess pricing policies.

Data and context parameters. The data that where gathered and analyzed during the study varied from dynamic consumption data, to static context variable measurements. Specifically, there were three different consumption measurement modes, within different subgroups of the users w.r.t. the frequency of measurements: measurements every 15 minutes, every 1 hour and every 10 seconds. Weather data were measured in a weekly basis. Also, historical and context data were taken into account: historical consumption measurements, users' demographic data, geo-spatial and qualitative (size, style, year built and construction material of house) information of the households.

Modeling. The designed model of the study had the following capabilities:

- Real-time intelligence and interaction for instrumented and interconnected cities

- Resource optimization and decision support for optimizing city's performance
- Behavior models and incentives for maximizing the impact of citizen engagements
- Integration of information about data, process, and people
- Provision of a common city view to multiple stakeholders
- Improved decision support for operations, policy planning, evaluation, and strategy

This way the following analysis capabilities were offered to the users:

- Information integration for smart meters (near real-time and historical), weather data, demographic and profile data, house data, geospatial information, etc.
- Data cleansing, extraction, transformation and loading from multiple sources
- Data models for energy, water, transit and transportation, and buildings
- Metrics computation for individual, utility, and city-wide/regional aggregation
- Business analytics and data mining algorithms for baseline creation, anomaly detection, forecasting, trending, and alert generation for individuals, utilities, and city planners
- Incentive design to support consumer behavioral change
- An engagement engine for optimal user experience management, social networking and collaborations, team formation, and reward computation
- A user interface with customizable widgets for visualization and interaction with information, insights, and incentives
- Dashboards for comparing and contrasting resource consumption performance by different user groups
- City-wide resource consumption business intelligence and analytics
- A collaboration platform for multiple stakeholders including city management, agencies, utilities, citizens· Resource consumption behavior models

Results and insights. The analysis results are divided in two categories: Consumption analysis results and Behavior analysis results. The most important are enumerated below:

Consumption results.

- The participants in the study managed to achieve an average reduction in consumption of 6.6% per household. If extrapolated to a full year, this means savings of ~12900 liters per household annually.
- Assuming that the 151 households are a fair sample of the City of Dubuque, the aggregate annual water savings across 23,000 households with smart meters in the city would be 245841 tons of water. This corresponds to savings of $190,936 a year in total.
- Pilot participants reported leaks at a rate of 8% compared to 0.98% of city-wide (i.e. 12 of the 151 pilot households, versus 226 out of 23,000 smart meter households). It is estimated that 30% of households on average have leaks.
- An active participation rate of 44% or 134 out of 303 users, including 35% (106) portal users, was achieved.
- Pilot households actively engaged in the experiment reduced water consumption the most –10% – compared to a group of 9000+ users with no portal access.

- Various portal functions assisted users in making sense of how they used water and appeared to have helped maintain their interest in water conservation with hourly usage, consumption comparison with other households (anonymously), participation in the weekly game (anonymously), daily news, and online chat being the five most important aspects cited.

Behavior results.

- In a survey (58% of the pilot participants that used the portal responded to the survey), 56 households that used the portal multiple times reported the most benefits and they include:
- 77% said that the Water Portal increased their understanding of their water use.
- 70% felt it helped them assess the impacts of the changes they had made.
- 48% felt that it helped them conserve water.
- 61% reported making a change to a water appliance or in the ways they used water (or both) during the study period (e.g., they took shorter showers; fixed leaks; purchased water efficient appliances, and altered watering system in the yard, etc.).
- 48% reported that they planned to make changes to their water equipment or ways of using water in the future.

3.4.1.3. Discussion

Conclusively, some general points extracted from the two aforementioned studies, w.r.t. to gathering, analyzing and exploiting time series consumption data, are the following: (a) It is important to take into account, apart from the consumption data themselves, all available contextual data and parameters that might affect consumption; (b) It is important to select the proper theoretic models, algorithms, features and processing models that best suit each sub-task of the problem; (c) User interaction and engagement, apart from enriching the information to be analyzed, is a factor itself for consumption behavior change; (d) Seasonality and demographic data are very important factors that affect consumption; (e) Gathering and analyzing data in several granularity levels is valuable for different sub-tasks of the problem, as well as for different stakeholders.

Although DAIAD aims at implementing a large scale, integrated framework for both users-consumers and water demand management experts and stakeholders, and perform studies of much larger scale, we believe that the aforementioned points constitute valuable guidelines towards our goals.

3.4.2. Data Analysis on Time Series

Analyzing water consumption data reduces to a problem of analyzing *time series of measurements*. Time series data, due to their nature, hinder the effectiveness and efficiency of classical machine learning and data mining algorithms. Thus, several of these methodologies and algorithms need fine-tuning, proper parameterization, adaptations and extensions in order to be applied on time series data. Although there have been several works (mainly pattern recognition and event detection) on analyzing time series in other domains, such as astronomy, remote sensors, heart monitoring, there is not much scientific work on water consumption data. Next, we first present some general background on algorithms for analyzing time series data, and then we briefly describe specific works we have identified on individual data domains.

The general idea of pattern recognition and event detection in time series data is to manage to identify a set of functions/curves/models that best represent a series of measurements for a period of time. Upon that, this information can be utilized in several tasks of the problem: produce aggregate statistics for the consumption of a user, find similar users w.r.t. their consumption curves, forecast future consumption behavior based on the trend of the consumption curve, find anomalous consumption behaviors based on significant divergences of measurements compared to the expected consumption, etc.

For example, some methods keep track of the mean and variance of the measured time series and, when a new observations' distance from their mean exceeds a certain threshold, they consider it an event. The calculation of the threshold is an issue, which can be dealt by giving more weight on the more recent measurements of the data (moving average – exponentially weighted moving average). Further, it is often the case that trends appear on the data as effects of various seasonal causes (e.g. day of the week, season). These effects can be taken into consideration and be incorporated into the model. For example, if the consumption measurements prove to increase every Friday (by observing historical measurements), we should not take this increase as an event, except if it exceeds a certain threshold. There is also the case of multivariate data, when the quantity we observe is possibly dependent on multiple variables, such as weather, demographics, etc. In that case too, the modeling of the consumption behavior should take into account the changes in the values of these context variables, when there is a divergence in the measurements compared to the expected values.

In the case of spatio-temporal data we have observations taken on various locations. The general approach there is to divide the space in regions and search for interesting events in each region. This strategy can also be applied time-wise: divide the measurements in fragments of time and exploit the fragments to identify common patterns. Finally an important notion is the effect of a feedback loop between the user of the system and the algorithm. This means that a user can ask a question and then give some feedback on the response (for example, classify some of the instances of the response). The system can then incorporate the new information to its models to improve its performance.

The Tutorial on Event Detection[42] presents a thorough overview of methods to detect events of interest in time series data.

3.4.2.1. Categorization of Algorithms

The work of [WS+12] contains a very comprehensive and timely survey on pattern recognition in time series. It identifies three distinct areas of research on the problem: (a) Algorithms for pattern recognition; (b) Issues on representation of time series data in order to cope with their vast volumes and achieve efficient processing; (c) Similarity measures for time series, namely, how to represent time series data in order to accurately discover similar patterns between them.

As far as algorithms are concerned, these are divided into three categories: *supervised*, when the user input is required to assign labels to a training dataset that is going to be used (as it is or by training a model) to suggest labels, thus identifying patterns, for new items-measurements; *unsupervised*, when the algorithm identifies patterns on its own and the user can then view and make decisions on these patterns; *semi-supervised*, which is an intermediate mode where both labeled and unlabeled data are utilized in the pattern recognition process. Another division is between *memory-based* and *model-based* algorithms. The former utilize the available training set as it is, by directly comparing its data with new measurements, while the latter train an intermediate model that is applied to each new measurement to identify patterns.

[42] http://videolectures.net/kdd09_neill_wong_ed/

A classical memory-based algorithm for classifying objects is k-NN (k Nearest Neighbor) which calculates the similarity of a new object with every object contained in the training set. The algorithm has been applied in [DT+08] to compare and evaluate the utilities of time series representations and similarity measures. The authors of [WK06] proposed a semi-supervised classification scheme on time series that is based on 1-NN. The algorithm first trains a classifier using all labeled data. It then classifies unlabeled data, and adds the most confidently classified unlabeled data into the training set. This process is repeated until some stopping criterion is reached.

Decision trees are machine learning tools that use a flow-chart-like tree structure, in which an internal node denotes a test on an attribute, a branch denotes the outcome on a test and a leaf node denotes the class label or class distribution. While decision trees are defined for real data, attempting to classify time series using the raw data would clearly be a mistake, since the high dimensionality and noise levels would result in poor accuracy. In an attempt to overcome this problem, Geurts [Ge01] suggested representing a time series as a Regression Tree (RT), and training the decision tree directly on this representation.

Support vector machines is another widely used classification model, where the data are mapped into a multidimensional feature space and the model to be trained corresponds to a hyper-plane in this space that separates-classifies the data. SVMs have been used for time series classification by several researchers. [CP08] used SVMs with dynamic time warping kernel for brain activity classification. [KM+09] extracted features from ECG (electrocardiogram) time series using statistical methods and signal analysis techniques, and classified the data using SVMs. [EH+02] proposed to use genetic algorithms to extract time series features, which are then classified using SVMs.

Classification approaches can also be utilized in the task of event/anomaly detection in time series. The general idea is to observe past time series measurements, build a model that represents the behavior of the time series and then compare future measurements to it. If the new measurements diverge significantly, it is probable that they correspond to an anomaly. In order to achieve this, Keogh et al combined a statistically sound scheme with an efficient combinatorial approach [KL02]. Their method is based on Markov chains and normalization. Markov chains are used to model the "normal" behavior, which is inferred from the previously observed data. The time- and space-efficiency of the algorithm comes from the use of a suffix tree as the main data structure. Each node of the suffix tree represents a pattern. The tree is annotated with a score obtained comparing the support of a pattern observed in the new data with the support recorded in the Markov model. Memory-based anomaly detection approaches have also been proposed, such as the one in [KL06], where time series discords are defined as subsequences of a longer time series that are maximally different from all the rest of the subsequences. They thus capture the sense of the most unusual subsequence within a time series. Time series discords are superlative anomaly detectors, able to detect subtle anomalies in diverse domains.

Further, specialized clustering algorithms (unsupervised learning) have been proposed for time series. The authors of [LV04] proposed an incremental and iterative version of the k-Means algorithm called i-kMeans. The algorithm works by leveraging off the multi-resolution property of wavelets, which mitigates the dilemma of initial centers selection for k-Means. Rodriguez et al [RG06] proposed a hierarchical clustering algorithm for time series data streams. The algorithm incrementally constructs a tree-like hierarchy of clusters, using a correlation-based dissimilarity measure between time series.

Finally, as far as similarity measures between time series are concerned, the most usual metric is the Euclidian distance. An enhancement of the metric involves Dynamic Time Wrapping where the series are realigned before applying the distance. There is also the LCSS method which tries to find the longest common subsequence within two time series. While those methods achieve very similar accuracy results, the simple Euclidian distance is computationally more efficient so it is favored. Another approach is that of structural similarity. In this case we are not looking for precisely the same sequence of observations rather than sequences that exhibit similar behavior overall. An efficient method in this approach is the Bag-of-Patterns Representation (BOP). In this approach there is a vocabulary of patterns and each series is characterized by how many times it contains each pattern. So, series that are constructed from the same patters can be considered similar. The main problem with this method is to define the vocabulary of patterns.

3.4.2.2. Pattern Recognition on Time Series

The work of [WS+12] presents an approach for pattern recognition in time series, called Symbolic Aggregate approximation (SAX). This method divides the axis of the observed variable (time series measurements) in separate ranges and maps each value to a range. Thus, it transforms the sequence of observed values to a sequence of symbols. This achieves compression of the data (without big increase in error, when used with a well-known classifier (1-NN) on several different datasets), which is very useful when the datasets are too big to fit in the available main memory. Then, by defining a sliding window of measurements, each sequence of symbols can be used as a pattern and count the appearances of each sequence in each time series. The more similar patterns two series contain the more similar they are. This method seems to capture the structural similarities in the presented test cases.

The work described in [SG+11] is concerned with the recognition of patterns in temporarily evolving multivariate data. A series of observations of multiple variables is available, as well as the assumption that they can be separated into periods of similar activity. The way similarity is defined is by means of the SVD decomposition of the data. Intuitively we suppose that two parts of the series are similar if they lie within the same hyperplane of the full space of the variables. The points are processed bottom-up and the sequent points that are found to lie in approximately the same hyperplane are merged together. In this way the series are separated into segments. Then similar segments are clustered together, again by the same metric of similarity. Finally re-occurring similar sequences in the original series are looked for. The novelty of this work is the use of the SVD decomposition as a similarity function.

3.4.2.3. Event Detection on Time Series

The authors of [GS99] aim at modeling phenomena whose behavior is not constant over time. Their method splits the total set of observations in intervals and models the behavior within each interval independently, by fitting a different function. The problem is called change-point detection problem. There are two different instances of this problem: one where all the observations of the phenomenon are available from the beginning and another where the observations arrive in real time. In all cases, when modeling a phenomenon using a function, there is an error involved, which expresses the difference between the model's predictions and the actual measurements. In the first case, where all the data are available, the approach searches them all serially to identify the point which, as change-point, minimizes the error of the model fitting. Then two intervals are defined, one up to that point and another from that point forth. The same algorithm is applied recursively on the new intervals for as long as there is a significant reduction in the model error. In the second case, where the data arrive in real time, the method keeps track of the last change-point and scans the data from there to the most recent observation to find a point which, as change-point, reduces the error of the model considerably.

[PP+09] focus on sensor-generated time series data. In their scenario, an event is identified by an interval of measurements that differs significantly from the underlying baseline noise. As an example, they consider the detection of micro-lensing events in astronomical data, during which more light than usually is detected, due to various astronomical causes. To achieve event detection, their approach first creates an estimate of the noise distribution. This can be done either analytically or, for more computational efficiency, by sampling of the available data. Then the algorithm scans the data using different starting points and interval lengths and calculates the probability that the data inside the interval where generated by the random baseline noise. If that probability is sufficiently small then it can be assumed that, at the given starting point and interval length, an event was observed. To scan through the data using different starting points and interval lengths, an heuristic algorithm is used, namely Powell's method. The algorithm developed at this paper was motivated from the field of astronomy and is proven to be very efficient at discovering events in astronomical data.

In [DD07] the authors describe a way of identifying malfunctioning sensors, by analyzing the stream of data they produce and subsequently removing the invalid data. Specifically, the system is developed for the case of temperature sensors but the principles can be used in other data domains as well. First they create a model for the measurement process. They assume that there is a baseline value for the temperature and a "departure" value which represents temporally local trends. Both those values depend on the day and time of day of the measurement and are learned using historic data from the sensors, which have been cleaned from effects of malfunctions by domain experts. This way, the system can make a prediction for the real temperature of any given time point. The value observed by the sensor depends on the real value for which an estimation is available, and the condition of the sensor. In the case of great variations from the estimated values it can be assumed that the sensor is malfunctioning. The proposed system achieves performance comparable to that of a domain expert.

3.4.3. Recommendations

A Recommendation System is a system that takes as input a set of users, a set of items and a set of past ratings of the users for the items and produces as output recommendations of items for the users. There are several algorithms, metrics and approaches that utilize item descriptions, user profiles, history and context to produce such recommendations. Recommender systems are classified into three main categories [AT05]: Collaborative, Content-Based and Hybrid Recommender Systems, will some categorizations include two more types: Demographic and Knowledge based:

- Collaborative [AT05, SFH+07]: The system generates recommendations using only information about rating profiles for different users. Collaborative systems locate peer users with a rating history similar to the current user and generate recommendations using this neighborhood.
- Content-based [AT05, MR99, PB07]: The system generates recommendations from two sources: the features associated with the items and the ratings that a user has given them. Content-based recommenders treat recommendation as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on item features.
- Hybrid [AT05, Bur07]: Hybrid recommenders combine features from different approaches to avoid the shortcomings of each one of them. The combination can be performed by: (a) weighting ratings from different approaches, (b) incorporating characteristics of one approach into the other or (c) constructing a unifying model that incorporates characteristics of different approaches. It has been empirically shown

[BS97, MMN02, Paz99, SN99] that hybrid recommender systems can provide better recommendations than pure approaches.

- Demographic [Bur07]: A demographic recommender provides recommendations based on a demographic profile of the user. Recommended products can be produced for different demographic groups, by combining the ratings of users in those groups.
- Knowledge-based [Bur00, Bur07]: A knowledge-based recommender suggests products based on inferences about a user's needs and preferences. This knowledge will sometimes contain explicit functional knowledge about how certain product features meet user needs.

Another categorization of Recommendation Systems is that of memory-based and model-based. Memory based systems apply heuristics to calculate rating predictions directly from the already rated items. On the contrary, model-based systems use the available ratings to train a model, which is then applied to produce ratings. An example of model-based approaches is k-Nearest-Neighbour (k-NN) algorithm that compares a target user's profile with the historical profiles of other users in order to find the top-k users who have similar tastes or interests. On the other hand, training a naive Bayesian model or an artificial neural network that can be used to calculate new ratings are examples of model-based approaches. In general, model-based approaches are more efficient and scalable than memory-based ones. However, they need either to function in an incremental way or to be periodically re-trained in order to be adaptable to changes (new items, change of user habits, etc.) [Mob07].

Tools and Libraries. The most recent effort for a complete, open source recommender system library is Apache Mahout[43]. Mahout is an ensemble of scalable machine learning algorithms offering various functionalities, such as recommendation, clustering and classification and data mining. It implements Collaborative Filtering, User and Item based recommendations and hybrid approaches, such as Item-based Collaborative Filtering and supports the development and integration of other recommendation algorithms. Mahout is implemented in java with emphasis on efficiency and scalability and can be deployed in a distributed computer framework (due to its Hadoop implementation), making it an ideal solution for applications that manage massive amounts of data. Other open source libraries that implement several recommendation algorithms are MyMediaLite[44] and SVDFeature[45].

Commercial applications. Some examples of commercial applications that incorporate recommendation functionality are listed below:

- Amazon.com is perhaps the most famous commerce site to deploy recommendations. Based on purchases and site activity, Amazon recommends books and other items likely to be of interest.
- Netflix similarly recommends DVDs that may be of interest, and has famously offered a $1,000,000 prize to researchers that could improve the quality of their recommendations.
- Dating sites like Líbímseti can even recommend people to people.
- Social networking sites like Facebook use variants on recommender techniques to identify people most likely to be an as-yet-unconnected friend.

In the context of DAIAD, recommendation needs to be adapted to the respective data handled. While classical recommendation usually revolves around real-world objects/concepts with explicit, fixed attributes, in our context, the items to be recommended are series of measurements. So, proper features representing these times series

[43] https://mahout.apache.org/
[44] http://www.mymedialite.net/
[45] http://svdfeature.apexlab.org/wiki/Main_Page

“objects” need to be extracted so that they can applied to recommendation algorithms. On top of that, proper modeling of the actual recommendation is required: in our case, the recommendation is not an existing item, rather than a change in user behavior that induces a change in the consumption measurements.

3.4.4. Personalization

Modern data collections and recordings of historic user interaction pave the way for personalized information retrieval. Personalized retrieval systems exploit user profiles and historic usage data to adjust, re-rank and filter retrieved data to serve individual information needs. Personalized retrieval aims at computing a ranking model for every user or groups of similar users. Different approaches including the impact of short- and long-term search histories [SH+04, ST+05], context [KC09, ST+05], query categories [TD+08, DS+08], search behavior and feedback [AB+06, Jo02, Jo06] or application of collaborative techniques to personalize search results on users with common interests/search intents [SH+04, GB+11] have been studied. Additionally, collaborative filtering techniques for personalized search [SH+04] and learning to rank-based approaches [ZC+07, QZ+07, CK07, AB+06, Jo02, Jo06, 26] also proved effective in many scenarios.

A prominent strand of research is based on exploiting historic user feedback. The impact of short-term versus long-term histories has been studied by [FI03, Jo02, Jo06] while [CF+06, DC+08] aim at capturing the context of the users, for instance by taking documents on the virtual desktop into account. The resulting models are essentially user profiles that are used to expand future queries and to refine the retrieved results. Many approaches incorporate state-of-the-art machine learning techniques to improve ranking results. [CX+06] study modifications of ranking support vector machines to reduce the error on top-ranks and to increase the importance of queries with only a few relevant documents in the training sample. In [QZ+07], the authors propose to learn multiple ranking functions for different ranks which are aggregated to induce the final ranking.

Finally, clustering methods are studied in combination with learning to rank strategies. [LW+07] propose to cluster results to discard probably redundant examples from a large training sample to render the resulting optimization feasible while [DC+08] cluster personalized ranking functions to group users for recommendation purposes. [BL+10] propose to learn multiple ranking models by clustering queries based on the topical information extracted by their results. They represent queries by aggregating feature vectors which are then clustered to obtain specific ranking models. The final ranking for new queries is being made by combining the models. Finally, the authors of [GB+11] propose to learn different ranking functions for different behavior and intents. These functions are independent of users or topics and rely on the latent search behaviors that are induced by several query types.

As far as DAIAD is concerned, the personalization goals align more with the methods described in the previous paragraph. Our aim is to first be able to identify personalized user consumption behaviors and then to exploit them to: (a) Find similar behaviors from other users, (b) produce personalized recommendations on consumption behavior changes, taking into account collaborative information from several users and (c) produce, when feasible, personalized consumption prediction models in the level of users or groups of users (e.g. similar users, neighborhoods, etc.). To this end, state of the art clustering and classification techniques will be adopted and extended to fit the needs of our setting.

3.4.5. Privacy Preservation

One of the most important tasks of DAIAD is the exploitation of crowdsourced consumption data and metadata to perform large-scale analytics on thousands of consumers. An important issue, though, is to be able to ensure the *anonymity* of every user, so that the engagement of users to share their consumption data is strengthened. At a first level, no actual namings will be gathered: users will be represented by identifiers that are irrelevant with any information that could give away their identity. Further, for metadata that could indirectly indicate the identity of users (location, demographics, etc.), state of the art anonymization techniques will be applied. Next, we provide a brief description of several techniques, covering two distinct data domains: relational and dynamic data.

3.4.5.1. Data Transformation Methods

Besides the particular algorithmic differences which are imposed by the nature of the data one has to tackle with, all approaches presented so far in the literature adopt a variation of one or more of the following data transformation methods.

Generalization. Generalization is the very first transformation method introduced in the study of data anonymization and, although it is rather general, it was originally presented along with k-anonymity. The principle of generalization is to substitute the original items that may lead to a privacy breach with generalized items, i.e. classes of items that have more than one instances, so that the initial rare items are hidden within frequent generalizations. A representative example of a generalization is the substitution of the items “goat milk” and “low-fat milk” with the more general item “milk”. Such generalizations may correspond either to classes of conceptually interrelated items, like in the previous example, or to classes which are arbitrarily defined by the anonymization expert. Apparently, a generalized item may be further generalized to another item forming, this way, hierarchies with more than one levels. In all of the existing approaches, the generalization hierarchies are tree-shaped, i.e., an item belongs to at most one generalization.

Suppression. Similarly to the case of generalization, suppression is another transformation method that was early introduced in the study of data anonymization. The basic idea behind suppression is that the rare combinations of items that may lead to a privacy breach are discarded; hence, they do not appear in the published dataset at all. In general, suppression naturally produces more information loss than the rest methods, but it can be very useful when the part of the data we want to suppress amounts to a negligible portion of the whole dataset. Thus, it is usually combined with other data transformation methods, more often with generalization and noise addition. In some works, the suppression of items is followed by the addition of new ones in their place. In the latter case, the added items may correspond to the average value of the suppressed ones (if we address items in a numerical domain), the centroid of a cluster (both for numerical and categorical items), etc.

Disassociation. Instead of generalizing or suppressing the terms in the published data, a disassociation-based approach reduces the structural information contained in the original dataset. The key idea is to hide the fact that certain items appear together in the same record. In this sense, some approaches aim at disassociating those items whose combinations can lead to a privacy breach. Regardless the specific implementations, all disassociation methods amount to breaking the records of the initial dataset into smaller ones which can then be published separately or in the form of reconstructed records after using a random permutation.

Noise Addition. Although thoroughly studied from a theoretical perspective, noise addition is a relatively new method in the field of data anonymization. It was initially presented in the context of statistical disclosure control, and more recently in the differential privacy paradigm [Dw06]. As its name declares, this method introduces noise in the data, usually Gaussian or Laplacian, with the aim to distort the initial records and at the same time preserve (as far as possible) the statistical properties of the original dataset. It can be applied to numerical microdata as well as to categorical data through appropriate ordering. Its advantage on the previous data transformation methods is that it does not depend on the background knowledge of the adversary; hence, it can cope with powerful adversaries who may also have statistical knowledge about the original dataset and not only instance knowledge. However, a significant disadvantage of the noise-based privacy-preserving approaches is that they are usually very strict; a feature that results in publishing a very small portion of the initial data (in some cases equal to publishing only the frequent records and suppressing the rest of the data).

3.4.5.2. Privacy on Relational Data

Privacy preservation was first studied in the relational context. In [Sw02] the author introduces k-anonymity (where the adversary is not able to link a record to less than k distinct entities) and uses generalization and suppression as the two basic tools for anonymizing a dataset. Incognito [LD+05] and Mondrian [LD+06] guarantee k-anonymity for a relational table by transforming the original data using global and local recoding respectively. A different approach is taken in [NC07], where the authors propose to use natural domain generalization hierarchies (as opposed to user-defined ones) to reduce information loss. Multirelational k-anonymity was proposed in [NC+07] and extends the notion of k-anonymity to multiple relations. The idea here is that information associated to a specific person is spread across many different relational tables which have several functional dependencies. The transformation of the data in this work relies on global recoding.

The aforementioned approaches protect the published data from identity disclosure i.e., they do not allow an attacker to associate a specific record to a specific person. Still, they do not prevent an attacker from linking some sensitive items (attributes) to a person. To address this problem the concept of ℓ-diversity [MG+06] was introduced. Anatomy [XT06] does not generalize or suppress the data, but instead it disassociates them by publishing them separately. A recent work presented in [LL+12] applies ℓ-diversity on a relational table by disassociating the combination of items that can lead to a privacy breach. The intuition behind this kind of splitting is to produce low information loss without compromising privacy. The respective algorithm, called Slicing, protects also from membership disclosure.

3.4.5.3. Privacy on Dynamic Data

The very first approach that takes into account the dynamic nature of a dataset was presented in [XT07]. In this work, the authors study the privacy breaches arising in the re-publication of a dataset that has been updated through insertions and/or deletions. Their approach applies k-anonymity in a dynamic setting and it is based on a variation of the generalization method, called counterfeited-generalization. In contrast to the offline anonymization algorithms, there are also some approaches focusing on real-time anonymization of data which are highly dynamic. A representative work of this case has been presented in [ZH+09]. It studies the problem of anonymizing streams of data in real time. The authors employ the k-anonymization paradigm by means of processing the statistics of the incoming streams on the fly in order to produce groups of indistinguishable streams. Other similar approaches that also employ k-anonymity for the instant anonymization of streaming data are those presented in [CC+08] and [LO+08]. The main idea behind these works is the use of an online clustering method that produces anonymous groups of streams, each one having a size equal to k.

4. Water Demand Management

Water suppliers invest substantial funds in the development and implementation of water demand management strategies to ensure that future water demand can be met. Typically, such strategy development consists of several steps, of which the first comprises the determination of the actual and past water demand structure. This includes answering the following question: *Who uses how much water under which circumstances and how have these figures changed in the past and for what reason*?

In order to answer this question, it is necessary to thoroughly analyze the actual water demand and its changes in the past with respect to its causal factors. This analysis will yield a list of factors, which influence the water demand to a greater or lesser extent. More importantly, some of these factors can be influenced directly or indirectly by a water supplier while others cannot. It is evident that the former group will become integral constituents of any demand management strategy. This is not to say, however, that the latter group is of no interest. Although these factors cannot be influenced by the water supplier, they are subject to changes in the basic conditions of water supply. Prominent examples of such changes are the economic and demographic development and climate change.

To the extent that these and other relevant changes can be anticipated, the second step of the strategy development sets in: *the extension of past and recent developments into the future*. Based on the identification of relevant water demand factors and their anticipated future changes a model can be built, which does not only show the *development* of future water demand under the assumed circumstances, but also the *type* and *strength* of influence of those factors that can be changed by the water supplying organization. The third step, eventually, leads to the development of the water demand management strategy in the narrow sense. It identifies and quantifies by means of the model those factors possibly enabling the reconciliation of water demand with the given water supply.

As the determining factors of water demand play such an eminent role in the development of the respective management strategy, it will be the aim of this section to present the state of the art of scientific knowledge about these factors and interventions addressing them in order to manage or change water demand. Thereby, the review will begin with presenting findings from literature on factors influencing water demand. The next subsection will review interventions in order to change water demand which we will distinguish according to the factors which are addressed. Finally, a third subsection will review the existing knowledge and lessons learnt from large scale trials, i.e. field experiments, in which water users are exposed to a more intense interaction with the water supplier aiming to interfere with his water consumption behavior.

4.1. Determinants of water demand

Water demand management is based on particularly two reasons: (1) in many countries or regions water is a scarce resource, and (2) water consumption is often connected with the use of energy and detergents as well as with the costs for waste water treatment. This implies that water cannot be supplied and used arbitrarily, but supply and use ought to be managed sustainably. In order to identify suitable measures for steering water demand, in particular economists have studied the parameters that influence water demand and, accordingly, can be addressed to manage the latter. Among the parameters are those, which can be used by a water supplier (e.g. a utility) to directly influence demand and those, which cannot. In the former group the price of water is the most prominent, but there are also other factors that can be addressed such as psychological constructs including *awareness, attitudes or norms*. In the latter group of factors, which cannot be influenced by the water supplier alone, households (*representing the typical client of a water utility*) as well as *geographical characteristics* are relevant. Knowledge about both groups is helpful, on the one hand to forecast water demand, on the other hand to manage or change water demand by addressing factors that can be influenced.

4.1.1. Price

While, in the long run, the use of water is determined by a wide variety of social, cultural and individual characteristics, the price of water seems to be the most important lever to influence the demand for water. How important it is, is shown by the debate about water as a "*human right*" and the obligation (of any government) to make *a certain quantity of water accessible to every person regardless of her ability to pay*.

Yet beyond this issue of basic need, price can be, and has been, used to manage water demand in many cases. This leads to the question of how *responsive* water users are with respect to changes of the water price. The basic economic concept for measuring this responsiveness is *price elasticity*, i.e. the percentage decrease of water demand brought about by a one percent increase in price. The price elasticity depends on a variety of factors that will be examined in detail below. Before this can be done, however, it has to be specified in the next section how the price increase comes about, how it is implemented in actual water tariffs and how these tariffs and their changes are perceived by the clients.

4.1.1.1. How does price matter?

There is an enduring debate among economists as to which price is relevant for determining price elasticity in the context of water demand management (see Arbues et al. [AG+03], Klein et al. [KK+07]). In theory, the *marginal price* of a good should be relevant for a buyer's decision to buy this good or not. This presupposes, however, that the client knows exactly what *the next unit she intends to use will cost*. In turn, this presupposes that she knows *exactly* her water *tariff* and the *quantity* of water she has used by the moment of her buying decision. At this point, it makes sense to investigate the existing types of tariffs and the challenges they may bring about.

As explicated by Arbues et al. [AG+03], designing a water tariff is a complex task as it seeks to reconcile such diverse objectives as *allocation efficiency, equity, sustainability and financial stability*, to mention just a few. To accommodate this set of objectives, the tariff commonly includes *fixed* and *variable* elements. The fixed elements reflect at least one part of the fixed cost of the water infrastructure and render the revenues less dependent on the quantity of water actually consumed. If the fixed element includes a certain amount of water subject to use without further cost, it can also serve as a *social aspect* of the tariff.

The variable element can be *uniform* (i.e. charging the same price for every (additional) unit of water used) or *variable* with the price for every additional volume of water increasing or decreasing continuously or stepwise after trespassing certain thresholds. The latter two are respectively called *increasing and decreasing-block rate tariffs*. Decreasing tariff schemes adequately reflect the low share of variable cost of water supply and are therefore considered economically efficient. On the other hand, they tend to promote the consumption of larger quantities of water. Increasing rate tariffs, by contrast, promote water conservation and are considered more equitable and explicitly redistributive [MC91]. However, they may adversely affect groups of people who, for health or other reasons, use greater amounts of water. In addition to these more regular elements, tariffs can comprise elements accommodating seasonal changes in water demand and supply and the distinction between peak and off-peak consumption.

Returning again to the initial question of which aspects of the tariff matter, what their impact on the responsiveness of water users is and how this responsiveness can be measured, the following answers can be found in the literature.

While it is economically reasonable to consider marginal prices as relevant in the case of uniform variable prices, a complication arises under block rate tariffs. Earlier studies used the marginal price corresponding to the block relevant for the last unit of consumption. Since Taylor [Ta75] and Nordin [No76] this simple approach is modified and a "difference variable" is introduced, which takes into account the difference between the water consumer's actual water bill and what she would have to pay if all units were charged the marginal price. Nordin [No76] argued that this variable should represent the income effect imposed by the tariff structure. In a variety of empirical tests, however, this hypothesis could not be confirmed [AB+03]. The reactions to this finding have been twofold. Some researchers argued that the estimation of this effect was done *incorrectly* and that ***the use of aggregated data*** was the major source of this failure (Schefter and David 1976). Yet a majority of researchers came to the conclusion that the difference variable is *irrelevant* because ***consumers lack the necessary detailed knowledge*** about the structure of their water tariff and, in fact, the difference variable amounts to a share of the household income too small to be relevant [NM89].[46] As a result of this debate, the average price was used in an increasing number of studies. Listing a large number of studies and their outcome in terms of price elasticity (see also the next section) Arbues et al. [AG+03] show that in many cases the choice of the variable (i.e. marginal or average price) does not seem to affect the results. If a difference is stated, demand tends to be more responsive to the average price.

It is evident that, in order to save water, people need to be *aware of their own water-using behavior* and be able to *classify* it as *lower or higher consumption* in the first place. If they are not aware of their consumption pattern, they will not be able to change it. Billings and Agthe [AB80] could show that indeed many water users ***are not aware of their consumption pattern*** or the price they pay for water. Yet the share of these ignorant water users was found to depend on certain factors. While Klein et al. [KK+07] and Neunteufel et al. [NR+10] report that *low-income households* exhibit a significantly stronger responsiveness to the water price than *higher-income households*, Renwick and Green [RG00] are able to quantify this effect. Especially in cross-country studies, however, it turned out that price responsiveness depends on the share of water expenditure from total household income [NR+10]. So, responsiveness can be high despite a high income level, if water prices are also high.

Increasing the price of water does not always lead to a change in the used quantity. A certain basic quantity of the water used in households for *drinking, cooking and various aspects of personal hygiene* (including sanitation and washing clothes) is considered to be *essentially insensible* to the price of water. It represents the minimum quantity satisfying the basic human need for water, to which every person should have access. This also constitutes the **human right to water (CESCR 2002)**. Martinez-Espineira and Nauges [MN04] approached this issue econometrically by means of a Stone-Geary utility function, which distinguishes between a price-sensitive and a non-price-sensitive demand component and allows for the quantification of both. For the city of Seville in Spain they found a price-insensitive quantity of **2.6m^3**/capita/month, which is **40%** of the total consumption (6.35 m^3/capita/month). For Germany, Schleich [Sc10] calculated a similar insensitive volume of **3m^3**/capita/month, which in this case represents

[46] In fact, water consumers do not only tend to be ignorant of most details of their water bill; they receive such bills only once a year in most cases. This challenges additionally the assumption that water users refer to the marginal price when deciding about their consumption level.

77% of the average total consumption. Both studies form a basis too small to draw a general conclusion, but they provide an idea as to the size of this price-insensitive component.

4.1.1.2. Differences in price elasticity

There is a large number of studies analyzing the determinants of residential water use and almost all of them come to the conclusion that *price has a significant influence on water demand*. Regardless of the price applied Grafton et al. [GK+09] found that the introduction of a *volumetric* water charge leads to a reduction in water consumption by **31.4%**. If the price is taken into account, in almost all studies, price elasticity[47] turns out to be *negative* and *rather weak* (in economic terms: *inelastic*). This means water demand falls with increasing price, but relatively it does not change as much as the price does. In their review, Klein et al. [KK+07] quote an average price elasticity of -0.49 [BB+02] and a range between -0.02 and -0.75 for 75% of the estimates (Espey et al. [EE+97]), both of which are consistent with a similar list compiled by [AG+03]. Grafton et al. (2009) [GK+09] arrived at a slightly lower value of -0.41 in their study. While this wide range of elasticities appears to reflect a certain lack of statistical reliability at first sight, the large degree of variability becomes more reasonable when understood as the *outcome of various influences*.

One factor influencing price elasticity is the *time* it takes until a *change in price translates into the respective change in demand*. It seems reasonable that the effects of measures taken to reduce water consumption is more limited in the *short run* than when water users are given more time to respond. Accordingly, long-run elasticity is expected to be stronger than short-run elasticity. This proposition is confirmed by a series of studies [DN+97][M087][NT03] showing that short-run price elasticity is in a range between -0.03 and -0.52 around an average of -0.2, whereas long-run elasticity ranges between -0.1 and -0.77 with an average of -0.5. In the study by Grafton et al. [GK+09], both short-run (-0.38) and long-run elasticity (-0.64) were found to be somewhat stronger, but their difference remained largely the same. The level difference can be interpreted in terms of other factors influencing the price responsivity. Yet from all studies it can be concluded that the *short-run elasticity is by about 0.3 lower than the long-run elasticity*.

Income is another factor influencing price elasticity.[48] Klein et al. [KK+07] and Neunteufel et al. [NR+10] report that low-income households exhibit a significantly stronger responsiveness to the water price than higher-income households. Quantifying this effect, Renwick and Green [RG00] show that households with an annual income of less than USD 20,000 were *five times* more responsive to a changing price than households with an income of USD 100,000 and more. There is no systematic analysis of this issue beyond this exemplary case. Additionally, it should be noted that not income alone may be decisive. Especially in cross-country studies it turned out that price responsiveness depends on the *share of water expenditure from total household income* [NR+10]. So, responsiveness can be high despite a high income level, if water prices are also high.

Above, in the discussion about the relevance of marginal or average price, block-rate tariffs represented a complication because, in this case, water consumption is determined by both the marginal price and a difference term. In the context of price elasticity, block prices again exhibit an influence, which is related to the latter complication. Especially in the case of an increasing block tariff, water users do not simply respond to the price of the last unit of water they consume; calculating their *opportunity cost* they also respond to the price of the lower

[47] Mathematically, price elasticity is defined as the ratio of change in demand quantity (in %) over change in price (in %). For normal goods it has a negative sign, because increasing price tends to give rise to a reduced demand.

[48] This effect is not to be confused with the direct effect of income on water consumption, which will be discussed in the next section.

block and the threshold between the two. As this opportunity cost is higher in the case of increasing block tariffs (as compared to uniform or decreasing block tariffs), it is not surprising that they could be shown to give rise to a stronger responsiveness [CH+02], yielding a 0.25 higher (in absolute terms) price elasticity (Dalhuisen et al. 2001).

Geographical location seems to matter a lot. The differences of price elasticity *between countries can be significant*. Yet this does not appear to be influenced by economic development or wealth. Also, as is shown by [KK+07], substantial differences with similar size occur within countries – between federal states or even cities – and even when controlling for household size and income. So, it remains unclear what the real explanatory factors behind these variations are.

Seasonality is reported to influence price elasticity in some cases. In countries where people are used to water their gardens in the dry season, this use of water appears to be less essential than other water uses. Accordingly, the price elasticity of the outdoor demand in the summer is estimated to be 5 to 10 times higher than in the winter [KK+07].

Eventually, there are artifacts that do not influence the actual price responsiveness of water consumers, but the econometric estimates of the related price elasticity. The nature of the data makes the difference in this case. Accurate price elasticity can only be estimated from *individual water use data*. Even household data are not expected to make a large difference. The situation changes, however, when more aggregate data from communities or even larger entities are used. In this case individual extremes are averaged out and, as a consequence, elasticity values appear to be smaller than they actually are. According to Dalhuisen et al. (2001), this effect can make price elasticity appear by 0.22 lower than it really is. A similar argument holds for *aggregation* along the time axis (i.e. using yearly instead of monthly, weekly or even daily data).

4.1.2. Household characteristics

Beside the price of water, certain characteristics of the water users and their households seem to be the most decisive determinants of water use quantities. In contrast to price, user-related factors *cannot be influenced by the water supplying company or community*. Most important from the economic perspective are the *income* of the water user and the existence of *alternative, less expensive water sources* like private wells. Other more physical structural determinants discussed in the literature are the *size* of the household, the *age* of respective household members and its *geographical* location. In the previous sentence, the reference made to households instead of individual water users points to a general phenomenon in the context of water use statistics: *for technical reasons* water metering, which forms the basic data source, is usually done on the *level of buildings or households*, but very *rarely on the level of individual users*. So, households are the most wide-spread and basic unit of water use.

4.1.2.1. Household income

In economic terms, water is a *normal good*, which means that water demand increases with increasing income. In accordance with this, the income elasticity of water demand is positive. As Dalhuisen et al. (2001) show in their meta-study, income elasticities show a mean of 0.46 and a median of 0.28, but the range of values is considerably smaller than for the price elasticities.

Like price elasticity, income elasticity depends on a variety of factors, which are to a large extent responsible for the reported variability. One of these factors is income itself. As Agthe and Billings [AB97], Saleth and Dinar [SD00] and Schleich and Hillenbrand [SH09] find in their studies, higher income households exhibit lower income elasticity.

Other significant factors identified by Dalhuisen et al. (2001) are the time perspective and the type of tariff system. In average, long run elasticity is by 0.34 smaller than short run elasticity, which can be explained by the habituation effect and, thus, the lower attention paid to income increases occurring over a longer period of time. With respect to tariff systems, the mean income elasticity under decreasing block prices is approximately 1.1 higher than in increasing block and uniform price schemes.

Eventually, like in the case of price elasticity, there are artifacts that do not influence the actual income responsiveness of water consumers, but the econometric estimates of the related elasticity. Specifically, the type of price variable used for the statistical analysis makes the difference. As Dalhuison et al. (2001) find out, income elasticities based on marginal prices are 0.27 higher than corresponding elasticities based on average prices.

Another methodological issue discussed in this context refers to attempts to identify the wealth of a household independent of the income of one or more of its members. As discussed in [AG+03], the value of the property appears to be a good proxy for household income – with the additional advantage that it could be assessed without asking the household members. Unfortunately, this relationship is not always confirmed in other studies. While living in a one-family house increases water consumption significantly in the study of Messner and Ansmann [MA07] for the city of Leipzig, Schleich and Hillenbrand [SH09] cannot confirm this effect for the entirety of German communities. Maybe, the latter failure to confirm is due to the aggregate nature of the data, which tends to "dilute" all effects and, thus, renders them less significant. Dalhuisen et al. (2001) confirm this effect by pointing out that aggregation of household data yields lower income elasticity.

4.1.2.2. Household size and age

Next to price and income, the size of the water consuming households has been assessed in a large number of studies on the determinants of water consumption (see overviews in [KK+07][NR+10]). In all these studies it is confirmed that the volume of water used *increases with the number of household members*, but the increase is less than proportional to the latter. Typically, the water volume is found to increase approximately by the square root of the number of family members [AV06][SH09].

In some studies, also the age of water users was assumed to influence water consumption. Typically, the age structure was assessed as the share of household members above a certain age. It turned out, however, that the results were often insignificant or not consistent between different studies. While Nauges and Thomas [NT00] found younger family members to use more water than older ones, Schleich and Hillenbrand [SH09] arrived at the opposite result. While in the former case, it was argued that youngsters might be less careful in using water and younger people might demand more frequent laundering than older ones, in the latter case older people were speculated to have more time to spend for outdoor activities such as gardening, which lead to higher water demand.

4.1.2.3. Education of household members

Higher education in general seems to have *little influence* on water consumption. While Grafton et al. [GK+09] identify a small but significant influence, Schleich and Hillenbrand [SH09] are unable to show a significant effect. While, in the latter case, the lack of significance may be due to the use of aggregated data, the effect is expected to be low even if more disaggregated data could be used. This expectation is in line with results for environmental behaviour in general (e.g. Homburg and Matthies, 2005).

4.1.2.4. Water saving technologies

Water saving technologies have been and are being employed in a significant share of households in various regions. Schleich and Hillenbrand [SH09] are convinced that this type of technical progress is the primary cause for the significant reduction in water use experienced in Germany between the 1980s and 2005. Examples of such water-saving innovations are *laundry machines* showing a decrease in water consumption from some 150 liters in 1980 to about 40 liters per wash in 2001 or dish washers with a decrease from about 50 to less than 15 liters over the same time period. *Two-flush and reduced-volume flush* toilets and *more efficient shower heads* are other means to cut down water use in two water-intense applications by up to one half [NR+10]. To determine the effect of these technologies on water demand on a *large scale is not so easy*, as this effect is masked by other effects such as increasing income-induced water use. Therefore, the number of studies investigating this water-saving effect is rather small. In a longitudinal study in Miami (Florida, USA), Lee et al. (2011) found a reduction by 11 to 15 percent after exchanging shower heads, toilets or clothes washers and even larger effects when several measures were combined. Herber et al. [HW+08] found a reduction by 15 percent for the low-volume flush toilets alone and another 14 percent for the use of highly water-efficient clothes and dish washers.

4.1.2.5. Private wells

Private wells are an alternative source of water, which can complement commercialized water supply wherever underground water is easily accessible. They are especially common in rural contexts, where they were the main water source historically and many uses (e.g. irrigation) do not require the water quality provided by the supply network. From this perspective, it is surprising that only few studies included private wells as potential demand factors in their investigations. To our knowledge, only Schleich and Hillenbrand [SH09] explicitly assessed the effect of wells on household water demand and found a significant but small effect, with the presence of a well leading to a reduction of drinking water consumption by 1.5 percent.

4.1.3. Geographical determinants

The variation of (per capita) consumption of publically supplied water across countries is quite large. In the **EU**, this volume ranges between **100 and 280 liters per capita and day**. The largest volumes are used in Sweden and Norway, where the supply is plenty and inexpensive. Interestingly, among the largest volumes are used in **Spain**, where water is not in plenty supply – but the price nevertheless rather low. In this context, Grafton et al. [GK+09] confirm what has been shown in the preceding part of this section: *pricing* (and, in the first place, charging) *of water matters*. Household characteristics such as size and income make a difference as well, but they do not differ so much across the EU. Especially the study of Willis et al. [WS+11] focuses on another cause of increased water consumption: arid climate and the need to cope with it, e.g. by irrigating gardens or take more frequent showers.

4.1.3.1. Weather

Willis et al. [WS+11] show by the examples of (parts of) the USA and Australia that under the conditions of hot and dry weather, low water price and absence of incentives to save water, water consumption is indeed higher than in more temperate regions. This changes however when the water price increases or other incentives to save water are provided. The recent development of water use figures in the south of Australia is an example for the latter effect. Another example is *Greece*, where per capita water consumption is only about *one half that of Spain*, although both are exposed to very similar climate. Apart from these more specific country comparisons, it is shown by Klein et al. [KK+07] that weather (or climate) exerts a significant effect indeed and the *lack of precipitation* is a better predictor for an *increase in water use than a higher temperature*. Temperature makes a difference only if it is really hot. In the case of precipitation, it is again not so much the amount of rain falling over a certain period of time that matters. People seem to adapt quite well to very different conditions without expressing significant changes in their behavior as long as the conditions appear regularly. The situation changes, however, when a *drought arises* and extends over a substantial period of time. In this case, people respond by spending more water in order to avoid harm or a loss of welfare. From this perspective, it comes as no surprise that Arbues et al. [AG+03] and Schleich and Hillenbrand [SH09] find that the *number of consecutive days without rainfall* is a much better determinant for water use than the *average rain fall over longer periods of time* (e.g. one month or year).

4.1.4. Psychological determinants

In addition to the economic, socio-demographic and geographical determinants of water consumption, as well as more psychological factors such as knowledge and awareness of the own water consumption pattern or the price of water, which have been discussed in the first part of this section, research in the field of environmental psychology has shown that several psychological factors influence environmentally relevant behavior such as water consuming behavior. Most research in the field of environmental psychology has focused on *residential energy-consuming behaviour and choice of transport mode*. However, studies on psychological factors influencing water consumption are rare. In the following, based on the general environmental literature, we outline the relationship between environmental awareness and corresponding behaviour, as well as more comprehensive models to explain behaviour by means of various determinants.

4.1.4.1. Environmental awareness and behaviour

Many studies were conducted over the last decades to investigate the determinants of different types of environmental behaviour from a psychological point of view. Repeatedly, it was *assumed that environmental behaviour is determined by and closely related to environmental awareness* [HK96].

In order to test whether water conservation is guided by environmental awareness, also named as *environmental concern*, Willis et al. [WS+11] designed a questionnaire allowing them to determine whether, on the one hand, the responding individuals are guided by environmental concern and, on the other hand, they apply water conserving behavior. On the basis of the survey outcome they were able to identify two clusters of individuals, those with very high concern (VHC) and those with only moderate to high concern (MHC). Eventually, they could show that the households with higher levels of environmental concern and a stronger attitude towards water conservation consume less water indeed. Interestingly (but not surprisingly) the conservation of water is restricted to those end uses that can be influenced behaviorally. In all other cases, both groups do not exhibit a difference with regard to water consumption behavior.

However, in general, empirical results have indicated only a weak relationship between general environmental awareness or attitudes and various types of environmental behavior [HM98][Ka96]. Different reasons are cited for this gap between attitudes and behaviour [HK96][HM98][Ka96], amongst others:

- **Lack of specificity of environmental behavior**. Often various types of environmental behaviour have been summarized into a global score of environmental behaviour, though environmental behaviour is not a homogeneous construct. Thus, an individual may hardly engage in all possible forms of environmental behaviour, but in specific types, e.g. due to the specific behavioural costs or the individual attitude towards the behaviour. For example, he may engage in recycling behaviour, but refrain from using public transport instead of driving car.

- **Different specificity of measurements of attitudes and of behavior**. Environmental awareness is often measured as a general attitude, whereas the behaviour is often assessed as a specific type of action. Closer relations are reported for specific attitudes towards a concrete behaviour. For example, attitudes towards using specific transport modes are more predictive for mode choice than general environmental awareness.
- **Different perspectives on environmentally relevant behavior**. The environmental assessment of behaviour is often not straightforward. A behaviour which is defined as environmentally friendly by the researcher is not necessarily perceived in the same way by the individual. For example, the purchase of milk or drinks in reusable bottles could be classified by the researcher as environmental friendly, while a survey respondent could disapprove it due to other ecological reasons such as increased energy use for the transport of glass bottles in comparison to carton packages.
- **Neglect of various motives influencing behavior**. Human behaviour and behavioural intentions usually do not aim to fulfill single motives, but have to be seen within a field of motives and purposes. For example, the choice of a specific transport mode may be influenced by the desire to travel in an environmentally friendly way; however, additional motives like safe, flexible and quick travel may also play a role.
- **Neglect of situational variables which influence behavior**. Further determinants such as behavioural control, awareness of environmental consequences of own behaviour, habits or further positive or negative consequences of behaviour have to be considered as determinants of behaviour. For example, if public transport is not available nearby home, the likelihood of choosing it is reduced.

Thus, the concept of environmental awareness proved to be *too holistic* to have significant explanatory power. Accordingly, psychological research showed that various and more specific variables better explain specific environmental behaviour. In order to integrate further and more specific variables which influence environmental behaviour, one important approach was to apply well established psychological theories of action to specific types of environmental behaviours [BM07].

4.1.4.2. Psychological theories of action

The action theories, which were most often applied to explain different environmental behaviours, are the theory of planned behavior (TPB; [Aj91]) and the norm-activation model (NAM; [Sc77][SH82]).

According to the TPB, behaviour is directly influenced by an individual's intention to perform the behaviour. Intention, in turn, is determined by (1) an individual's *attitude* towards the behaviour, defined as an overall evaluation of its possible consequences, (2) *subjective norms*, referring to the perceived expectations of other important persons, e.g., family, peers, neighbours (we will speak of social norms in the following), and (3) the *perceived behavioural control* (PBC), defined as a person's perceived ability to perform the behaviour due to non-motivational factors such as availability of opportunities and resources. The attitude towards the behaviour is conceptualized by Ajzen [Aj91] (cf. also Fishbein & Ajzen, 2010) as an expectancy-value model. According to this model, the expectancy that a specific behaviour results in particular consequences and their evaluation, i.e., the valence of these consequences, are assumed to determine the overall evaluation of the behaviour.

Studies using the NAM explain behaviour as being influenced by (1) a *personal norm* to engage in the specific behaviour, denoting a strong intrinsic feeling of obligation. Prerequisites of the formation and activation of this personal norm are (2) the *awareness* of a related problem that needs to be solved, (3) the *awareness* or identification of the specific behaviour as an effective action that contributes to mitigating the specific problem (we will speak of response efficacy in the following, according to Lam and Chen [LC06]) and (4) the *recognition* of the personal ability to engage in these actions which may correspond very well to the TPB's PBC. Besides personal norms, the consideration of (5) *social implications*, i.e., a perceived social norm, as well as (6) *non-moral implications* of action influence the behaviour. These influences are also included within the TPB by the concepts of subjective norm and the attitude concept. A further influential variable contained in the NAM approach is (7) the *assumption* of *responsibility* for one's own actions and their consequences.

Values which are a factor also discussed to influence environmental behaviour are not an explicit determinant within these two psychological theories of action. Values are central, but rather distant determinants of human behavior. They influence and thus are mediated by variables such as attitudes and norms which represent more direct and more specific determinants of behaviour. Consequently, these specific determinants have more explanatory power to explain specific behaviour and are common components of psychological action theories. Values, in contrast, are useful concepts for holistic lifestyle concepts which are applicable to various fields of behaviour.

For both theoretical frameworks, the TPB and the NAM, substantial empirical evidence has been collected for a variety of behaviours[49], such as environmental behaviours (for the TPB, e.g., [HH07][KG03][KP+99][TP+04]; for the NAM, e.g., [GF+03][HN91][HB+01][Th99]). More recently, various researchers proposed to integrate both concepts into one model [BM07][Ma05].

To summarize, according to psychological studies on various environmental behaviours, various psychological factors, such as certain *personal attitudes* or *personal or social norms* might *increase or decrease water consumption*. With regard to the sustainable use of water, it is necessary to be aware of the problems, i.e. negative consequences of water consumption and link them to the own behaviour. Based on these basic conditions, ***consumers have to know and be aware of effective behavioral options to reduce water consumption***, so that a personal norm to perform these behavioural options can develop and be activated in relevant situations. If these options are perceived to have overall positive consequences and if consumers perceive own abilities and opportunities to conduct the specific behaviour, likelihood increases that water saving behaviours are implemented. These factors and finally behaviour is also influenced by relevant others, i.e. by social norms, values and perceived behaviours of others. For the field of water consuming behaviours, studies on the specific relevance and role of these factors are still rare, more research is necessary to deepen and confirm the outlined relations for this area.

4.2. Interventions to change water demand behavior

Literature on effects of interventions to change water demand behavior is rare, to our knowledge. In this chapter, we will therefore present basic options to change water demand behavior. With regard to empirical findings on the effects of these interventions, we will mainly draw on studies in the energy field and present literature for the water domain where available.

---

[49] Another prominent area where the TPB was proved to be a feasible predictor is the topic of health behaviour (cf. Armitage & Connor, 2001); the NAM was actually developed – and effectively applied – to explain altruistic behaviour (cf. Schwartz, 1977; Schwartz & Howard, 1982).

4.2.1. Interventions focusing on structural determinants

4.2.1.1. Changing the price

From the economic perspective, changing the price of water would be the primary instrument for changing the demand for water. This is also in line with the sustainability perspective, if, as is often the case, the actual price does not cover the cost – including the external cost – of water supply. From the social perspective, higher water prices are often considered as contradicting the "*human right to water*". According to this, at least the quantity of water needed to meet the *basic human needs should be affordable for every person*. If this basic quantity is indeed supplied at a low price, it is still possible to meet sustainability criteria by increasing the price for the quantity of water going beyond the basic needs. However, the resulting price scheme – increasing block price – is not reflecting the actual cost structure of water supply with its *high fixed and low marginal costs*. From the economic perspective, it is therefore considered as rather inefficient. Beside used quantities, price schemes can also refer to other parameters such as water availability. Seasonal price increases, for instance, can reflect the higher scarcity of water in months with lower precipitation.

With regard to effects of such price schemes, Schleich and Klobasa [SK13] evaluated the effects of residential energy prices differentiating between peak and off-peak times (with a ratio of peak to off peak prices of 177%) in a large field experiment in Germany. The experiment with more than 1,500 households lasted for six months. Results suggest this time of use pricing led to average percentage reductions in peak-demand of 6% to7 %, while off-peak demand was neither reduced nor increased. Thus, also the total demand was reduced as reaction to the pricing scheme. These results are in line with findings from most time of use pricing experiments in other regions (mainly in North-America). Differences over the time of the experiment (6 months) were not observed. As this time frame is rather short, the authors regard behavioral changes as relevant for the observed changes in demand while investments in energy-saving measures might play a role in longer periods. Thus, in the long term, the effects might even be larger, if households do not return to long-term habits after a certain time.

4.2.1.2. Regulatory instruments

The access to water can also be restricted by means of regulation. In this case, water use for irrigation, car washing (i.e. non-essential uses), for instance, could be prohibited. Or water supply can be limited to certain times of the day. In case of probitions, changes of behavior depend on further factors such as if compliance of water users is or can be checked and if violation is sanctioned, or if the prohibition is comprehensible to the users and accepted by them. In case of limitation of water supply, users would have to adjust to these limitations and might shift at least part of their demand to times with no limitation. As is shown by Michelsen et al. (1999), the net effect of such restrictions is still as intended: less water is used altogether.

4.2.1.3. Technical approaches

Instead of changing water use pattern by means of price changes or regulating the supply side, interventions can also refer to the demand side. Employing technical devices that enhance the efficiency of water use, for instance, will reduce the actual water demand without compromising the comfort of water users. However, somebody would have to pay for the necessary investment. Often this is not done by the water user without further interventions although the investment would pay off within a relatively short period of time. In this case, it may make sense for the water supplier to invest into a retrofit program for more efficient shower heads and so on, if reducing the demand by such a program is less costly than extending the supply accordingly.

4.2.2. Interventions focusing on psychological determinants

Besides the factors addressed by the interventions in the previous section, also psychological factors such as norms, attitudes, beliefs, and knowledge have been shown to influence behaviour and can thus, be addressed in order to change behaviour. As mentioned above, empirical studies on interventions to reduce residential consumption behaviour have been conducted mainly for energy consumption, so that we will draw on literature from this field in the following. For the transfer of findings to the field of water consumption, differences as well as similarities between consumption behaviours in both fields have to be taken into account.

4.2.2.1. Information

In a review of psychological literature on interventions to reduce energy consumption, Abrahamse et al. [AS+05] found that most studies focused on such psychological variables, but did not include changes of structural or situational determinants of the behaviour. With regard to basic psychological determinants of behaviour such as problem awareness, awareness of consequences of the own behaviour as well as awareness of behavioural alternatives and sufficient competences to conduct the behaviour, in general interventions conveying information are feasible [Ma05]. When considering and comparing the effect of information provision in different studies, it has to be taken into account that the specific content of information can be very different (e.g. general information on a problem vs. detailed information on how to act to contribute to a solution of a problem) and thus, information can have an effect via affecting different determinants of behaviour. Moreover, information can be provided in very different ways which can be relevant for its effectiveness.

Delmas et al. (2013) offer the most comprehensive meta-analysis of studies on information based interventions promoting energy conservation. On average, the reductions of electricity consumption achieved by the various strategies ranged at 7.4%. However, information on monetary savings or on monetary incentives (payment or rate changes) led to increased energy consumption instead of helping to reduce consumption. An explanation given by the authors is the "*licensing effect*", i.e. by the information, users may *learn that cost and/or potential savings are small, and they are entitled to use energy as they are paying for it*. These findings indicate that a strong *focus on pricing information and strategies might not be as effective* as often assumed by stakeholders and practitioners.

4.2.2.2. Combination of measures

Often information is required as pre-condition, but information alone is usually not sufficient to induce action. According to the review of Abrahamse et al. [AS+05], information measures led to more knowledge, but did not always lead to behavioural changes or reduction of energy consumption. Rewards led to a reduction; however, they did not have persistent effects. Feedback measures were effective as long as *feedback was given in a consistent and frequent way*. Especially, the combination of feedback with other measures, e.g. *comparisons with other users* and a *competition* with awards as incentives was evaluated as successful by the authors.

Similarly, in a study of Abrahamse et al. [AS+07] a combination of tailored information, goal setting and tailored feedback has proven successful to reduce residential energy consumption. In their study, also comparative feedback was used, but did not show an effect on energy consumption. Their results are mainly consistent with other studies. With regard to comparative feedback, mixed findings are reported in literature (cf. discussion in [AS+07]). Explanations for the absent effect of social norms could be that the comparative feedback was not given immediately following the behavior in question, that the reference group might not be relevant for participants, and that social norms might not be very salient as there was no communication with members of the reference group. As the authors point out, more research is needed on why social influences seem relevant in some cases, but not in others.

4.2.2.3. Different forms of feedback

With regard to effects of different forms of feedback, Darby [Da06] conducted a review on findings in literature studying feedback on energy consumption: *Direct feedback* relates to immediate feedback from the meter (if it is clearly visible) or an associated display monitor. Its effects range from 5 to 15% according to [Da06]. Findings indicate that users with high consumption may respond more to direct feedback than users with low consumption. This feedback type is suitable to give the consumer adequate information on different end-uses in a simple way, by showing consumption when an appliance is switched on vs. when it is not in use. [Da06] points out that every household should ideally be able to notice current consumption and its changes without having to switch on other optional feedback devices. The term *indirect feedback* denotes feedback that has been processed in some way before reaching the energy user, usually via billing. Reductions of consumptions induced by this feedback type range from 0 to 10% dependent on context as well as on the frequency and quality of information given. E.g. accurate, frequent billing will provide better insight into variation of consumption and causal factors than single billing once per year. With regard to the provision of *comparative information*, comparing with previous recorded periods of consumption appears to be more effective than comparing with other households or with a target figure.

According to [Da06], long-term effects of feedback are supported if it successfully supports psychological determinants for water saving behaviour, development of new habits and investment in efficient appliances and technology. Thereby, additional information, normative measures or advice on behavioural saving options or information on efficient technology seems useful. With regard to incentives given for a certain time-frame in combination with feedback, effects are likely to fade away when incentive is not given any more. Generally, continued feedback is necessary for enduring effects.

With regard to feedback via smart metering, Gölz et al. [GG+12] and Klobasa et al. [SK+12] studied the effects of feedback via energy smart metering on consumption of households in Germany and Austria. Households could choose if they would like to receive feedback via an internet portal or a postal letter with each option chosen by approximately 50% in both studies. Feedback was combined with advice on energy saving measures. In both studies, the effects on energy consumption ranged on average around reductions of 4% (*without an effect of the different feedback options*). In the study of [SK+12], the largest reductions were achieved in households with a medium level of consumption while households with a very high or very low consumption level hardly showed changes of demand. Different effects of measures on different consumer types were also suggested by the reviews of [AS+05] and [Da06].

Gölz et al [GG+12] also studied evaluation of the online and postal feedback instruments, which they offered to the participants of their study, and they assessed the way the internet portal was used. Feedback was mainly evaluated as positive and as informative, useful, understandable and user friendly. Participants of the study stated that the feedback raised interest on the topic of energy consumption. 15% of the participants felt under pressure to reduce their consumption by receiving feedback. About 25% of the sample was afraid of problems regarding privacy protection. The online information portal was mainly used in the first months after implementation. The type of information which was used most frequently were values for hours and days and advice for energy saving. In the second months, rate of use dropped drastically (on average by 50%), while in the following time a continuous decline of usage was observed.

On the whole, the presented literature outline shows that a *combination* of different measures ***tailored*** to the needs of the target group and the context is most effective to change behaviour. General information, for example information on a problem (e.g. consequences of water consumption), relevance of own behaviour and behavioural options are often necessary but not sufficient to change behaviour. *Feedback* can enhance the effectiveness of provided information. Thereby, a clear and visible feedback seems to be a necessary component to allow users to learn details about their consumption and to help them to identify specific options to reduce their consumption. An immediate direct feedback in combination with frequent and detailed indirect feedback which provides detailed information on effects of behavioural changes seems effective.

4.3. Large-scale trials and their implications for water demand management

4.3.1. Overview of water metering across Europe

Water metering throughout Europe is not ubiquitous in nature but varied from country to country; this trend is evident even within individual countries. The degree to which metering has been adopted is highly fragmented for a multitude of reasons. In the EU, the *Water Framework Directive*[50] is based on the idea that modern water management needs to take account of *ecological, economic and social functions* throughout the entire river basin [Eu14]. There has been a trend towards higher water prices in real terms throughout Europe over the past 20 years and there are wide variations that exist in individual countries across Europe. This is due a wide range of factors that determine water prices and the costs associated with recovery [Eu14]. With a greater environmental awareness largely helped by directives such as the Water Framework Directive, it can be seen that large scale water metering alongside water efficiency has begun to play a more frontal role as a solution.

The current widespread metering situation across Europe is largely dependent on the historical organisation of how countries typically have been providing water. Environmental pressure has been the driver to reduce overall consumption in this situation, in particular the threat of *water scarcity*; this is commonly due to climatic variations such as those experienced in Mediterranean countries. In recent years the predicted impacts of population growth coupled with climate change and emerging socio-economic trends [BM06] have also been a driver towards the case for metering. However, legislation, policy and water utility structures also play a role on the degree to which uptake occurs. These imminent threats have spurred countries on to change the way they approach water demand management in its entirety.

Across Europe there is a wide spectrum of countries with successful metering penetration. For example, currently in England and Wales 46% of customers have a water meter (see Table 1), whereas in France there is a far higher level of metering as standard practice. The typical type of metering across Europe can be classed as '*dumb metering*' which just records the volume of water passed through and has to be manually read in order to bill for consumption. The water meters to date have typically been *accumulation meters, pulse meters or interval meters* [Ma10]. A meter is typically deemed as the *fairest* way to pay for water. Further, pricing and tariff structure of water and how this can be applied (i.e. seasonality, peak timing and the effect of this) can only be understood through large scale metering roll-outs to see whether there is a behavioural change due to cost implications. This is where the application of

[50] http://ec.europa.eu/environment/water/water-framework/

AMR, intelligent and smart-metering come into play, as they are able to assist with providing timely feedback and information which can assist water utilities in making correct water demand management choices.

The following sections look across Europe at the current metering situation in order to understand the key motives behind metering trials and programmes, alongside the effectiveness of the large scale trials.

4.3.2. Examples of large scale trials

4.3.2.1. United Kingdom

In England the current water services provider structure is privatized, which in turn results of customers having *no choice or influence* over which company provides them as it is determined by location of the customer and company. Privatisation occurred in 1989 as part of the Thatcher regime and has since remained that way [Of12]. Water utilities used to be a nationalised system. This privatisation has had implications for metering and currently there is *no legislation* that companies must abide by to meter consumers; politicians have shown no will to mandate it, and most recently any metering policy has been left completely out of the Water Bill [UW14].

The Government has a aspirations to reduce the daily per capita consumption down from 150 litres/person/day [St10] to 130 litres/person/day [De08]. Water companies are obliged to report their business plans to Ofwat (the water regulator) in order to monitor progress and see that they adhere to what they propose and that it aligns with Governmental aims and best interests for consumers. Through these plans it is predicted that by 2015 half of all homes in England and Wales will be metered [De11].

Consumers have the option to pay on a *fixed tariff* scheme, which is not variable, opposed to paying for what they use, which is the function of a basic meter. Therefore they are *not assuming financial responsibility* for the amount of water they use. The current percentage of customers that are metered is an average of 46% (Utility Week, 2014). Table 1 shows the percentage of customers that are metered across all water providers in 2013/14.

Company	% of water customers metered	Company	% of water customers metered
Affinity Central	44	South East	59
Affinity East	77	South West	76
Affinity South East	93	South Staffordshire	31
Anglian	75	Southern	70
Bristol	42	Sutton and East Surrey	43
Cambridge	67	Thames	33
Dee Valley	55	United Utilities	36
Essex and Suffolk	55	Welsh Water	37
Northumbrian	29	Wessex	56
Portsmouth	23	Yorkshire	46
Sembcorp Bournemouth	64	All companies	46
Severn Trent	39		

Table 2: Sourced from http://www.utilityweek.co.uk/news/a-smart-move-for-water/975152#.U8RLjvldXA4

There is however an effort by the UK Government to try and build a consensus around legislating compulsory metering [St10] as it is deemed the 'fairest way to pay'. The case for metering has been made by the Fairness on Tap report[51]. Due to the nature of water companies in the UK it is up to them to work out the associated costs and benefits of increasing the amount of water metering [De11]. Due to these complexities, a *blanket approach* to metering across the country has not been taken, as the government believes water companies are best placed to find the appropriate local solution through discussion with their customers [De11].

There is also a high level of concern from Ofwat and the Environment Agency about the financial implications metering could have on heavy water users, such as large families who could also be on a low income (The Walker Review, 2009; [St10]). This is despite the fact that on average UK households with a water meter tend to use 10% less water [GW+10], and a better level of leakage detection.

From 2013 companies will also be able to support households through the implementation of a well targeted social tariff [De11], which takes into account other aspects (such as income) to adapt the price of water accordingly. In light of this, it is expected by the Government that water companies who choose to roll out metering programmes, will handle this transition very carefully and take a bottom-up approach through explaining the purpose of the metering programme and ensuring that there are transitional tariffs in place to cushion the transfer to the new system [De11]. A transitional tariff is beneficial both for the water company and the customer. It offers the customer an opportunity to find out if they are paying more through having a meter, with the option to revert back to being non-metered if it is not of financial benefit (a strategy undertaken by Southern Water).

In the past years water utilities have begun to initiate widespread metering, but following many different paths due to the individual nature of its company. Metering has mostly been driven by water stress; already on a peak day in London demand outstrips supply [UW14]. It has also been found that on average Thames customers use a third more water today that they did 30 years ago (Thames, WWW). The most common route taken has been to install 'dumb meters'. This can be seen in the case of Anglian Water, who at 75%, has a high meter penetration level but has been installing meters since 2005, while since 2010 has installed 100,000 one way radio meters [UW14].

More recently, there has been a shift towards the use of AMR meters through some of the larger water companies. On the whole this technology has not been universally adopted. In the UK there is a strong emphasis on proving the business case development which has been applied to water scarcity, and it is predicted that the market is to move straight towards smart metering (Frost and Sullivan, 2011). This can be exemplified by Thames Waters launch of a new 15-year project to begin a smart meter roll out starting in Bexley (Thames Water, 2014; Utility Week, 2014). The decision to begin such a large smart metering trial was through the demonstration of the broader business case over the past years (Utility Week, 2014).

4.3.2.1.1. Southern Water Case Study

One of the most ambitious metering trials and subsequent roll outs in the UK has been conducted by Southern Water. Southern Water supplies 555 million liters of drinking water to 1 million households every day [SW14] and is one of the largest suppliers in the UK. They are located in the South East of England which is an area of high water stress and has suffered droughts, with subsequent Temporary use Bans during the 2012 drought [Wa13]. This is predicted to get worse, with current trends in the UK seeing a shift in population towards the South. This, coupled

[51] http://www.waterwise.org.uk/resources.php/7/fairness-on-tap

with natural growth and consumers appetite for a more water intensive lifestyle, is only going to exacerbate the stress on water resources in this region.

Southern Water is more than halfway through its AMR programme to universally roll out AMR meters. It is replacing 100,000 existing dumb units and installing 478,000 first time meters. They are currently on track with just under 500,000 meters installed so far and 230,000 customers moved off rate-able value based charges (Utility Week, 2014).

Picture 42: Southern Water supply and wastewater treatment areas (Source: http://www.southernwater.co.uk/about-us/about-southern-water/our-business/)

4.3.2.1.2. Isle of Wight Trial

Purpose and aims of trial. As a result of being in a water stressed area, Southern Water began to embark on a universal metering programme in order to help customers reduce their consumption, which in turn would reduce the threat of supply not meeting demand. The aim of their metering programme was to ascertain the annual installation rate of metering in order to set the case for how many meters they would be able to install by 2015 [Ar11]. There was more to this than just learning about annual installation rates. Southern Water looked at how customers would *react* to metering. They also explored what they could do in terms of assisting the householder with making changes to their household water using appliances and behaviour to prepare them. They also looked at how they could accommodate the householder and reassure them that they would not be financially worse off through having a meter, and so looked to develop a *social tariff* as part of their long term aim to universally meter customers. As part of this aim, an initial trial was undertaken on the Isle of Wight, this was widespread [SW09].

Results. The national trial on the Isle of Wight consisted of 48,000 meters being installed over the duration of three years. This demonstrated to Southern Water than an annual installation rate of 16,000 meters/year was achievable [SW09], and that it could further increase as it is relative to the size of the team conducting the installations. One of the biggest insights was the *amount of leakage discovered* and so alongside the metering trial there was a policy of repairing and replacing defective supply pipes which contributed to a substantial and noticeable reduction in demand, especially in summer months [SW09]. A baseline consumption level had been established and it was reported that the metering had an impact of reducing the distribution input on the island by 12%. This number was derived from a baseline pre-metering base established in 1988 (Southern Water, 2009). It is through this trial that Southern Water have predicted that they will be able to achieve the target of installing 500,000 meters by 2015 which will result in 92% of their customers being metered and able to take control of their bills [Ar11].

Even though this trial is relatively small compared to Southern Waters long-term metering aims, there were positive results alongside lessons learnt about best practices when embarking on a large scale trial and metering implementation.

As mentioned, a 12% input reduction was noted following the Isle of Wight trial, but it was also noted that in Hampshire the distribution input increased by 3%, thus reinforcing the positive impact the metering trial had [SW09]. There are direct benefits associated with reducing the input revolving around (a) the costs and carbon associated with the generation and distribution of drinking water, and (b) the amount of pumping and treating of wastewater that would have been required should the metering implementation not have occurred. This has positive environmental benefits that relate directly to the water stress of the area, as through a reduction in consumption leads to a reduction in abstraction from rivers.

Aside from the environmental benefits experienced in the Isle of Wight trial, there were also lessons to be learnt and drawn upon about the importance of the relationship between the customer and utility provider. With the customer being on a meter opposed to a tariff structure, it enables the customer to have *autonomy* and ability to *directly influence bills*. In the UK there is a drive to increase housing available, which is likely to put pressure on existing supply and infrastructure. Therefore, by allowing customers to take more power over their consumption, habits will allow for growth without further exacerbating the supply issue in a stressed area [SW09]. Further, reducing resources used is beneficial to the existing infrastructure and distribution systems, as they would be able to cope with future population growth and subsequent demand.

4.3.2.1.3. Southern Water AMR metering trial

Introduction. Southern Water began a large scale metering roll out in 2010 with the aim to replace 100,000 existing dumb meters and install 500,000 AMR meters instead. For this, Southern Water selected Arad's Gladiator water meters which is integrated with Arad's 3G system. These meters allow Southern Water to remotely read individual consumption through drive-by technology [Ar11]. These readings are transmitted to a central database which is directly linked to the Southern Waters billing system [Ar11]. The AMR devices can log up to 4,000 reads with programmable intervals, and they can be downloaded remotely in less than one minute [Ar11]. There are time saving advantages through this choice of meter, as the AMR system will enable 20,000 readings to be taken a day, opposed to the 200 that could be taken manually when dumb meters have been used. Further, there are receivers installed in the company's AMR vans, which are linked to the navigation system, thus optimizing the route for efficient coverage (saving extra time, fuel, and carbon emitted) [Ar11].

Purpose and aims of the trial. Following the success of the Isle of Wight trial, Southern Waters aimed to install 500,000 AMR meters by 2015 to achieve a 92% level of metering penetration. The purpose was twofold. There was the motive of improving water demand management through a higher resolution and subsequent understanding of the amount of water customers used, as well as being able to drive a reduction in household consumption through bringing in a meter payment system. There was also the motive to reach out to consumers and understand what works and the best way to promote and integrate a mass metering trial. It was a case of running a trial at the same time as initiating a roll out scheme.

Participants and approach. Due to its universal nature, the roll out scheme was targeting all Southern Water customers throughout the supply area and so were not distinguishing between consumers. The trial has initially begun in the most water stressed areas of; Southampton in Hampshire, Horsham in West Sussex and the Medway area in Kent [Ar11]. The roll out started in 2010 and is set over a five year period. The roll out is phased in order to allow time for constant feedback, evaluation and adjustment as is required to make it successful.

Southern Water is very committed to providing good customer service and creating a positive relationship with their customers. This is reflective in the amount of consumer engagement there was throughout the trial and roll out. Prior to the meter installation Southern Water initiated a wide scale awareness campaign. This involved advertising in newspapers, radio and poster campaigning with the message that the South East is an area under water stress. The other message offered the customers the opportunity to 'save water, save energy and save money' through having a meter [De11]. Through a consultation process, Southern Water held meeting with local councils and community organisations including; Age UK, Citizens Advice Bureau and Neighbourhood watch [De11]. Customers received lots of written information about the new meter and how the billing process works. In addition to this, they are offering lots of advice and support to demonstrate to customers how they can make the most of the opportunity to reduce their water and energy bills to yield the greatest benefit from the scheme [Cr11].

A unique aspect of this roll out is the investment in a *Mobile Exhibition Unity*, which is to be parked locally whilst meters are being installed. This offers the customers the opportunity to come and tacitly engage with Southern Water advisors and receive demonstrations [Cr11]. It was also realized that the type of messaging needed to be different to attract different consumer groups due to the relatively *low cost of water* compared to the likes of energy. The economic arm was not always the best message, and so the *environmental and social norm ideals* were utilised to persuade customers that metering would be a good step forward (Southern Water Contact, 2014).

For the customers that were worried about being financially worse off through metering, Southern Water offered a transitional tariff. This tariff allowed customers flexibility through the first year of the meter being installed, with the option to have the meter removed if they were to end up paying more for their water. Southern Water agreed a rateable tariff with Ofwat where all customers can opt for a changeover tariff which can take them from rateable values to metered charges over a period of three years. Six months after the switch over, customers receive their first metered bill which will be the first indication as to whether or not they are saving money by being metered [De11]. Customers were also provided with feedback through the color bill they were using so they could have a '*month-on-month*' indicator as to whether they were consuming more or less water. If they consumed less, the bill was green but if they consumed more, the bill was purple [De11].

Results. As the AMR roll out begun in 2010 and is not set to finish until 2015, the results are yet to be fully understood. To date the trial has a total of 420,756 meters installed [SW14]. One of the most important results is the benefit of the *leakage alert system*. With the AMR meter system a leak can be detected in the customers' supply pipes and within their households. The alarm can detect leaks with a starting flow of only 1 litre/house [Ar11] and so even the smallest of leaks can be detected. It is predicted by Southern Water that the leakage picked up on customers pipes will save up to five million litres per day.

Summary. It can be seen that Southern Water has approached their metering trial with a very *grassroots* approach that is sympathetic and receptive to customers. Southern Water worked methodically to prove the business case for the wide scale metering alongside the environmental benefits it would bring. From the increased awareness of leakage, this will pay dividends for Southern Waters water demand management in the long run.

4.3.2.2. Malta

4.3.2.2.1. Water Services Corporation Case Study

Overview. The Water Services Corporation WSC is the water supplier throughout the Mediterranean country and archipelago consisting mainly of Malta, Gozo and Comino. Due to the island nature of the Maltese Islands, this presents a great challenge with water supply and distribution, alongside climatic charecteristics. The average rainfall in Malta is around 350 mm/year [MS02][EM05] with low levels of groundwater (around 40 mm^3/year) and no surface water (*albeit tiny freshwater pools which are ecologically important*). This makes water demand management a strategically and demanding conundrum to resolve.

The WSC is a governmental national water utility that is responsible for the management of the water resources. In order to try and meet the supply needs of the country, alternative methods have been looked for. This has resulted in around 19,000,000m^3 of water coming from desalinated sources in 2003 – 2004 [WS04], meaning that 50% of total water provided is from a desalinated source [Ca14]. Desalination is a resource-exhaustive process and with energy being *inextricably linked*, there is concern about how this will impact on the cost of water with a potential for out-pricing it in the future [Wa12]. The Government in Malta also sets the price for electricity, which in turn determines the cost of the water from a desalinated source (Malta Takes Control). So the concern and threat of *affordability of water* could become an imminent reality. Adding to the concern, it has been found that saltwater is infiltrating into the aquifers which supply around 60% of the country's freshwater. This could lead to a further shift forward to increasing the production of desalinated water [Go10].

A further method that has been explored in Malta is the *reverse-osmosis* of saltwater. This currently accounts for roughly a third of Malta's water from a total of three plants. Through this process, the cost of the water produced by the WSC is mostly from electricity, accounting for 75% of the price [Go10]. When Enemalta (the electricity provider) increased electricity rates for the large commercial and domestic customers, this also effectively increasead the rates for water by as much as 25% [Go10]. It can already been seen in Malta that there is a point approaching where people must make the tough decision between conserving water and electricity or paying more [Go10]. Malta have been *proactive* and there has been work surrounding developing technologies to get more water from each watt consumed in reverse osmosis plants. The energy used to desalinate 1 million litres of water dropped from 5 KWh (25 years ago) to 2.8 KWh recently [Go10]. However, 25% of total water is unaccounted for [WS04] which could be due to leakage, or due to utility theft which is a situation that has been seen in both the energy and water sector before in Malta [Go10]. In light of this and a need to install a method that enabled them to take control of their water resources, the WSC embarked on a metering trial.

Introduction. The WSC began a smart metering initiative with IBM in order to address the gap between supply and demand and to begin to safeguard consumers against future spikes in price as a result of a higher volume of water being processed through desalination and reverse osmosis. There has been a range of technology trialed such as; walk by and fixed networks alongside RAMAR and Zigbee. For the trials the WSC undertook four pilot projects, two in Gozo (2003 – 2004) and a further two in Malta (2005 – 2006). The purpose and aims of the trials was to assess which technologies worked successfully as well as providing consumers with the ability to understand the amount of water they use and how it is priced [Pa14]. Further to these initial trials, it was decided to run a trial in 2009 with the local energy company as a smart grid approach, called Integrated Utilities Business Systems (IBUS) [Ca14].

Approach and results. Throughout the IBUS project the WSC installed just over 202,000 transmitters out of a total of 255,000 needed for customers, which equates to around an 80% of households with the new smart meter installed.

This has resulted in 126,000 customers being billed through remote reading technology. This was a difficult number to reach as there were social problems encountered when trying to access the premises and install the radio frequency transmitters [Ca14] and so achieving such a high percentage of coverage is highly satisfactory result. Even though it was a successful and wide reaching metering trial and roll out, there were technological issues encountered. There were problems with the AMM system, such as a potential loss of communication. However when this occurred there was no data lost, just temporarily unavailable.

Despite these problems, there were tangible benefits to the WSC. It was found that they largely revolved around billing and the cash-flow as a result of this. With the ease of radio frequency systems, the WSC was able to *increase the billing cycle* as it is not as labour intensive to find out consumption levels for household [Pa14]. Further, having accurate access to billing data reduced the number of billing disputes due to the total elimination of estimated bills. This should help to repair the relationship with the WSC and customers and gives them a better customer service and communication platform.

From an environmental and water demand perspective, the metering trial has been of the upmost benefit with regards to the level of detailed information that is being provided for balancing real water loss alongside having better control of leakage detection [Pa14].

It is anticipated that the social issues around the metering will no longer be a future issue. Firstly, this is because the meter reading does not disturb customers and so less interference per se would be welcomed. On the Maltese islands there has been a history of tampering with utilities; through the AMM project this will be easier to detect and brings it full circle back to having a positive effect on cash-flow through people paying their bills instead of finding ways to not [Pa14].

This initial trial with IBM was not a total success as after the trial the WSC made several changes to the system implemented. A new contract was awarded to Suez Environment and Lyonnaise des Eaux through their joint subsidiary Ondeo Systems [Sa14]. This contract aims to equip the 400,000 inhabitants of the island [Sa14]. As part of this the initial short range AMR units were *removed and switched to 100% long range Ondeo Systems* AMR Units [Ca14]. Work is currently being conducted to install AMR meters of which 80% is completed, equating to 300 zones permanently being metered and logged so far. There are also now only 260 receivers for 316 km² [Ca14].

Summary. It can be seen that this is a different approach taken to implementing a large scale trial of water metering with the first trial not being unsuccessful per se. It is also an example of a trial that has been from a top down mandated approach and encountered forms of social resistance. However with an 80% level of widespread metering across the islands, it has been for the most part successful. The driver for this was not necessarily the environmental stress, but also the economic aspect through protecting customers in the future from high and unaffordable spikes in prices which could further exacerbate the tendency to not pay. The proactive method of initiating a widespread smart metering trial and roll out bodes well for the future of water demand management in Malta.

4.3.2.3. France

In France water services are considered as a public service which means that they must adhere to a certain number of characteristics including; equal access for all consumers, continuity of supply/quantity/quality, and adaptability to technical innovation [Re06]. The water law of 1992 states that '*water is part of the common heritage of the nation and that its use is for all*' [Sm06].

The way this service is provided is different through municipalities, as the local communities decide whether to directly manage their water services or delegate it to a private company through contractual agreements [Re06]. In contrast to the model in the UK, there is no centralised authority in charge of regulating the water industry in France [Re06] which is an important factor when communities are deciding on supply methods. The water pricing is based on the principles that '*water must pay for water*' and so subsidies should not be seen as the major source of financing [Se06].

This has led to a *progressively growing private sector*; it now has about 80% of the market [Re06]. In the case of private management, the relationship between the municipality and the private company can take different forms such as management contracts and lease contract. The lease contract is the most common form and is usually awarded for 7 – 12 years. The private firm is then responsible for the operation and maintenance of the water infrastructure but is not obliged to invest in it [Re06]. Water metering throughout France is already widespread and is the most common method to pay for water. Individual properties are metered and it is common in multiple apartment buildings to have a single meter installed. The cost per apartment is then worked out based on the square meterage of the property [Ba11]. Individual water metering is promoted by law and recently became mandatory for all new buildings in order to enhance early repairs of leaking taps or toilets. More radically, the new water law is promoting rainwater harvesting to replace the use of drinking water [Sm06].

The two main water distributors in France is Ondeo – Suez group and Veolia water [HL12]. With the size and competition within the private sector, France is already developing and trialing smart metering for water. This is largely due to the drive to retain customers, but also due to the nature of these organizations. They do not solely provide or deal with water, but also have specialized departments and business units focused on the development of smart metering technologies.

4.3.2.3.1. Mylhouse, Lyonaisse des Eaux case study

Introduction. Mulhouse is located in East France and is in close proximity to the Swiss and German borders. Mulhouse is made up of 13 municipalities and within the metropolitan region there is a population of 194,000 people [Ly13]. Its water authority (overseeing water distribution in the 13 municipalities), chose to install the remote reading solution offered by Lyonnaise des Eaux by the end of 2015 (awarded in 2013). Suez – Lyonnaise des Eaux is a spin-off company of Suez Environment, created to boost the business development of smart metering and form technological solutions [Ca14]. Through this installation, Lyonnaise des Eaux will have over 1 million smart water meters installed [Ly13].

Purpose and aims of trial. The city of Mulhouse is participating is a 'smart city' programme, which includes innovations to improve the quality of life for its residents [Ly13]. A contract worth 3.5 million euros has been awarded to install a smart metering solution offered by Lyonnaise des Eaux to all subscribers in Mulhouse by the end of 2015.

There are many aims to this project that are based around both the consumer and water services provider. The aims mainly revolve around water demand management and the more effective supply of water for residents in the future. On the consumer facing side, it will allow municipal authorities to help Mulhouse residents to manage their water consumption on a daily basis and enable them to be alerted if there is a suspected leak in their households. For the water providers the bill will be based on actual rather than estimated consumption, which is beneficial for cash-flow and account balancing. One of the biggest benefits of the system will be the ability for the city to monitor in real-time the efficiency of its drinking water network [Ly13].

Participants and duration. As part of this trial a total 16,000 people will have the remote water meter reading solution installed and have access to the system to look and monitor their daily water consumption.

Results. The trial is in its embryonic stages with installations currently taking place. The system is not anticipated to go live until 2015.

Summary. Even though smart metering technology for water is a relatively new concept, it is a logical step in France given the early adoption that was taken to metering as a standard method to pay for the water used. Through the nature of the Suez Environment group and relevant subsidiaries, it is interesting to see the technological advancements that are being made.

4.3.2.3.2. Biarritz, Lyonaisse des Eaux case study

Introduction. Biarritz is a city on the Bay of Biscay located in South West France, and is a popular tourist destination. As a result, its population can increase from 20,000 (off-season) to 100,000 people in peak holiday times [Ca14]. Alongside the population change there is a fluctuating occupancy in the properties, which is usually a short term rental given the seasonality and turnover of tourists. In order to accommodate this variable population, a seasonal tariff has been implemented, which runs from April through to October and is set at 15% above the normal rate for water [Ca14]. Through the traditional dumb metering method it is difficult to charge according to the tariff without it being a manual effort and labour-intensive process. Therefore losses can be made in financial terms. As a result it was decided that an AMR system should be installed [WD14]. The AMR system was telemetry-based to which the communication was limited to nine receivers of 169 MHz covering the entire territory of the municipality [WD14].

Purpose and aims of trial. The aim of the trial was to roll out AMR meters firstly in municipal facilities and subsequently in domestic properties in order to be able to apply the seasonal tariff.

Participants and duration. The first phase of the trial began in 2006 with the aim to integrate AMR for municipal facilities. The second phase of the trial began in 2011 and involved all the customers which equated to a total of 20,000 meters being installed [Ca14]. The requirement was that the deployment should take less than a year [WD14].

Results. The initial results of the AMR rollout were promising and largely successful. The aim of being able to charge non-residents for water on the seasonal tariff was no longer a problem in Biarritz. This provided the water authorities with a higher degree of control over domestic backflows and had positive implications for billing [Ca14][WD14]. Having access to their daily consumption enabled consumers to decide their own thresholds and take a detailed level of control over their consumption.

A further positive outcome of the trial was the amount of leakage that was found. In the first phase of the trial in municipal facilities there were quick and outstanding returns through leakage detection at places such as beach showers and irrigations systems. This success was also echoed in the domestic roll out with leakage suspicion communicated to 15% of owners in 2013 [Ca14].

From a water demand perspective, having the relay of live telemetry data enabled the operators of Lyonnaise des Eaux to measure the daily performance of the network and adjust its operation to changes in its performance. This provided quicker feedback than had historically been available, as the results were available immediately opposed to semi-annual reports. Between 2010 – 2012 Lyonnaise des Eaux helped save the equivalent consumption of a million inhabitants [WD14].

Summary. It was a positive trial with positive feedback from those who were receptive to smart metering; more than 100 enrolled immediately for the leak alert service available online or from a smartphone (The Water Debate, 2014) which shows good consumer engagement. The trial shows that through smart metering a seasonal tariff can be easily applied and yield good results.

4.3.2.4. Non European Trials

4.3.2.4.1. Dubuque, Iowa, USA

Introduction. In September 2009 the city of Dubuque announced a partnership with IBM in order to make Dubuque the first smart city in the US. The city has a population of around 60,000 [IB11] with an average water demand of eight million gallons daily [Da14]. The water in Dubuque is supplied by the Dubuque Water Department. The aim of the Water Department is to maintain a water distribution system that is inline and recognized by the American Water Works Association for efficient management practices [Da14].

The city of Dubuque has successfully undergone a community wide water meter replacement project whereby existing water meters were replaced with new meters manufactured by the Neptune Technology Group. These meters are connected via radio frequency through which the meter collects usage and remotely transmits this information to data collectors. The meters are able to detect; leaks, reverse flow situations and zero usage data on an hourly frequency. This project successfully ended in June 2014 [Da14].

Dubuque began a smarter water pilot study as part of the Smart Sustainable Dubuque project in conjunction with the community wide meter replacement project [Da14]. This trial was conducted with IBM.

Purpose and aims of trial. The purpose and aims of the smarter water trial was to build on the community-wide metering programme and to begin to embark on a campaign to become a Smart City in the US. For the consumers the aim of the trial was to provide households with the information, analysis and social competition around water consumption in order to test the hypothesis that 'informed and incentivised citizens would be able to conserve water more efficiently' [IB11].

A further aim of the installation of smart meters was to establish a new baseline for water consumption in order to help educate citizens but also give them a tangible target about how much to reduce their consumption. By being provided with access to their water consumption data, there was a goal to enable the city and its residents to be able to visualise and understand their patterns, as well as seeing the contribution this would make in their wider sustainability footprint. In order to portray this to the householder, they had access to statistics and displays on; gallons of water used, cost ($), pounds of CO^2, a display with daily or hourly consumption, historical water usage, comparative consumption over weeks and similar households [IB11].

The project included an aspect of social competition and gamification through social challenges. The online portal included games there were aimed at promoting sustainable behaviour and online forums to 'chat' with other users and exchange ideas about how they were saving water.

Participants and duration. The smarter water pilot study involved two phases of equal sizes; the total study involved 300 households for three months. An online portal was created which enabled households to view their consumption and other indicators related to usage.

The first phase lasted for of 8 weeks and involved 151 of the participants. Participants were given access to; the portal, weekly reports and support personal. They were provided with full access to all aspects of the system. At the same time there was a control group of 152 participants which had no access to the portal. The purpose of the control group was to allow a direct comparison to be made between the groups which helped to establish cause and effect.

The second phase of the project lasted for six weeks and involved all of the participants of the control group, which were given access to the portal with the same functionality as that of the first group. The purpose was to establish how both groups used the portal, but also make a direct comparison of the consumption of the control group before and after they had access to the portal [IB11].

Approach. The approach to this pilot was voluntary. Participants were recruited through the likes of the City of Dubuque website to sign up. Throughout the project there was an element of how social challenge can be employed to facilitate a large scale behavioural change across the households. This project was very socially driven through the online chat form, and being able to post messages on a pilot wide bulletin. There was deliberately not a lot of influence from either IBM or the City of Dubuque throughout the trial stages of the project. A local NGO facilitated with participants through; a weekly newsletter, training sessions, town hall meetings and an informal sustainability newsletter. The approach taken was bottom-up with the participants largely stimulating the interest and social interaction to generate continued interest throughout the duration [IB11].

Results. The results from the project were good with a noticeable reduction in household consumption. The pilot households who were the most actively engaged with the project reduced consumption by 10% when compared to a group of 9,000 who has no portal access. This shows that when consumers are able to get a better understanding of their consumption patterns, alongside social challenges and competitions, it can promote sustainable behaviour change [IB11].

Households were so acutely aware of their consumption patterns that there was a substantial increase in the rate of leaks reported. The city-wide average of 0.98%, increased dramatically to 8% from participants in the trials. As a result of the trial, a proposal was developed to extend the programme to allow 4,000 more households to have access to the portal and allow current users to continue to have access [IB11].

Summary. Even though the City of Dubuque case study is non-EU, it is a valid study as it is one that brings in the element of social interaction and gaming as part of initiating a change in lifestyle to live more sustainably. It also demonstrates the degree to which consumers can begin and want to understand their consumption, and how this can not only have positive impacts on the amount of water they use, but also with regards to understanding their consumption and being able to identify a leak.

4.3.3. Future implications on metering and water demand management

4.3.3.1.1. Trends

From the pan-European perspective and wider case studies, it can be seen that the overarching trend in water metering is a *shift away* from traditional dumb metering methods, towards AMR, radio frequency solutions and most recently, integrated smart technology solutions. According to research carried out by Frost and Sullivan [FS11], it is anticipated that government will not be the large driver in the case of the smart water metering market, unlike the electricity and energy market. Instead, the growth will largely come from utilities looking to take advantage of the level of network and operational efficiencies and associated savings that can be made.

If this is the case, then the level and flexibility of organizations will be seen at different rates across the EU. This is already apparent when current metering situations are compared (e.g. in UK and France). In France, the nature of the organisations such as Suez – Lyonnaise des Eaux, which has devoted teams to the development of smart metering, has paid dividends in the number of roll outs.

4.3.3.1.2. Drivers

The degree to which there is a need for innovation and alternative solution is down to situational drivers which vary across countries and regions. There are different reasons for initiating metering or more advanced smart metering solutions, but they mostly revolve around providing a *holistic solution* to water demand management. From the case studies that have been outlined it can be seen that the main drivers are as follows:

- Area is in current or predicted water stress
- Population growth and increased consumer consumption
- Easier implementation of seasonal or peak tariffs
- Informed management for current water supplies
- Increase in cash-flows and accurate revenue
- Reduce the amount of leakage and non-revenue water
- Prevent theft of resources
- Provide consumers with the tools to lead a more sustainable lifestyle
- An aspiration to be a ‘smart city’ with a better quality of life
- Ageing infrastructure and pressure on existing infrastructure
- Reduced energy use in the water life-cycle

4.3.3.1.3. Prospects

Although the outlook for water metering and smart metering in particular looks like a clear-cut solution, there are many practical aspects that need to be considered in order to ensure that a roll out is successful and useful. There is a real risk of consumer privacy breaches [GW+10]; An understanding is required about *who* has access to what levels of information. For example, when water is not being used, the house is likely to be vacant; if excess consumption is occurring at a time of water stress or drought with restrictions, how would a consumer feel to be told they need to use less water. In contrast with the energy sector there is no mandate in the water sector and without this and a common focus, water companies have *no shared vision* for smart metering. What the SBWWI Smart Metering Suppliers Forum would like to see in the future is common solution for a metering system [UW14]. It is expected that the global smart water market will continue to exponentially grow and that by 2020 the penetration of smart water meters across Europe will be at 50% [WS13].

5. References

[AB+05] Arroyo, E., Bonanni, L., and Selker, T. Waterbot: Exploring Feedback & Persuasive Techniques at the Sink. SIGCHI Conference on Human Factors in Computing Systems, p. 631-639, 2005.

[AB+06] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In Proceedings of the 29th annual international ACM SIGIR conference, pp. 19-26, 2006.

[AB80] Agthe, D.E., Billings, R.B., 1980, Dynamic models of residential water demand. Water Resources Research 16 (3), 476–480.

[AB97] Agthe, D. E., Billings, R. B., 1997, Equity and conservation pricing policy for a government-run water utility. Journal of Water Supply Research and Technology 46 (5): 252–260

[AG+03] Arbues, F., Garcia-Valinas, M.A., Martinez-Espineira, R., 2003. Estimation of residential water demand: a state-of-the-art review. Journal of Socio-Economics 32: 81–102

[Aj91] Ajzen, I. (1991). The theory of planned behaviour. Organizational Behaviour and Human Decision Processes, 50, 179–211.

[Al11] Allcott, H. Social Norms and Energy Conservation. Journal of Public Economics, 95(9-10), p-1082-95, 2011.

[Al82] Allen, Chris T. Self-Perception Based Strategies for Stimulating Energy Conservation. Journal of Consumer Research, 8(4), p. 381-390, 1982.

[AM+94] Aitken, C., McMahon, T., Waring, W., Finlayson, B. Residential Water Use: Predicting and Reducing Consumption. Journal of Applied Social Psychology, 24(2), p136-158, 1994.

[AM10] Allcott, H., Mullainathan, S. Behavior and Energy Policy. Science, 327, p.1024-1025, 2010.

[Am13] AMPHIRO. Hot water is energy | Amphiro. Retrieved April 9, 2013, from http://www.amphiro.com/en/products.

[Ar11] Arad (2011) Arad Case Study: Southern Water's AMR Implementation. Arad. Available online: http://arad.co.il/assets/CaseStudy_20111.pdf.

[AS+05] Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. A review of intervention studies aimed at household energy conservation. Journal of Environmental Psychology, 25, p. 273-291, 2005.

[AS+07] Abrahamse, W., Steg, L., Vlek C., and Rothengatter, T. The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. Journal of Environmental Psychology, 27(4), p. 265-276, 2007.

[AT05] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734–749, 2005.

[AV06] Arbues, F., Villanua, I., 2006, Potential for pricing policies in water resource management: estimation of urban residential water demand in Zaragoza, Spain. Urban Studies 43 (13): 2421–2442.

[Ba11] Barraque, B. (2011) Is individual metering socially sustainable? The case of multi-family housing in France. Water Alternatives 4 (2):223–224.

[BA11] Bengtsson, F., and Ågerfalk, P. J. Information technology as a change actant in sustainability innovation: Insights from Uppsala, Journal of Strategic Information Systems, 20(1), p. 96-112, 2011.

[BB+02] Brookshire, D.S., Burness, H.S., Chermak, J.M. and Krause, K. Western Urban Water Demand. Natural Resources Journal, 2(4), p. 873-898, 2002.

[BC+99] Bechtel R.B., Corral-Verdugo V., and Pinheiro J.Q. Environmental belief systems: United States, Brazil, and Mexico. Journal of Crosscultural Psychology, 30(1), p. 122–128, 1999.

[BG06] Backlund, S., Gyllenswärd, M., Gustafsson, A., Hjelm Ilstedt, S., Mazé, R. STATIC! The Aesthetics of Energy in Everyday Things.

[BJ10] Björkskog, C., Jacucci, G., Gamberini, L., Nieminen, T., Mikkola, T., Torstensson, C., Bertoncini, M. EnergyLife:Pervasive Energy Awareness for Households. Proceedings of UbiComp, p.361-362, 2010.

[BL+10] J. Bian, X. Li, F.-Li. Liu, Z. Zheng, and H. Zha. Ranking Specialization for Web Search: A Divide-and-Conquer Approach by Using Topical RankSVM. In Proceedings of the 19th international conference on World wide web, pp. 131-140, 2010.

[BL99] Brandon, G., and Lewis, A. Reducing Household Energy Consumption: A Qualitative and Quantitative Field Study. Journal of Environmental Psychology, 19(1), p. 75-85, 1999.

[BM06] Butler, D and Memon, F A. (2006) Water Demand Management. IWA Publishing.

[BM07] Bamberg, S., & Möser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. Journal of Environmental Psychology, 27, 14–25.

[Br01] Bredahl, L. Determinants of Consumer Attitudes and Purchase Intentions With Regard to Genetically Modified Food – Results of a Cross-National Survey, in Journal of Consumer Policy (24:1), pp. 23–61, 2001.

[BS+13] Beal, C., Stewart, R., Fielding, K. A novel mixed method smart metering approach to reconciling difference between perceived and actual residential end use water consumption. Journal of Cleaner Production, 60, p.116-128, 2013.

[BS03] Bamberg, S., Schmidt, P., Incentives, Morality, Or Habit?: Predicting Students' Car use for University Routes With the Models of Ajzen, Schwartz and Triandis. Environment & Behavior, 35(2), p.264-285, 2003.

[BS97] M. Balabanovic and Y. Shoham. Fab: Content-based, collaborative recommendation. Communications of the ACM, 40(3), 66-72, 1997.

[BT06] Bénabou, R. and Tirole, J. Incentives and prosocial behavior. American Economic Review, 96(5):1652–1678, 2006.

[Bu00] R. Burke. Knowledge-Based Recommender Systems. Encyclopedia of Library and Information Systems, 175–186, 2000.

[Bu07] R. Burke Hybrid Web Recommender Systems. In The Adaptive Web, 377 – 408, 2007.

[Bu11] BusinessGreen. (August 26, 2011). Europe ready to open tap on smart water meter market - 26 Aug 2011 - News from BusinessGreen. In BusinessGreen - news, comment and analysis for the low carbon economy. Retrieved April 9, 2013, from

http://www.businessgreen.com/bg/news/2104657/europe-ready-tap-smart-water-meter-market.

[Ca14] Carton, P. (2014) Benefits from the Latest Generation Automated Meter Management Systems. IWA European Utility Conference. Oslo 14-16 May. Presentation. Available at: http://cdn2.iwahq.info/all/EUC2014/9%20Business%20Cases%20in%20Smart%20Management%20of%20Water%20systems/2%20Benefits%20from%20the%20Latest%20Generation%20Automated%20Meter%20Management%20Systems%20-%20Philippe%20Carton%20%5BCompatibility%20Mode%5D.pdf

[CB+02] Corral-Verdugo, V., Bechtel, R., Fraijo-Sing, B. Environmental beliefs and water conservation: An empirical study. Environmental Psychology, 23(3), p. 247–257, 2002.

[CC+08] J. Cao, B. Carminati, E. Ferrari, and K.-L. Tan. Castle: A delay-constrained scheme forks-anonymizing data streams. In ICDE, pages 1376-1378, 2008.

[CD10] De Craemer, K., & Deconinck, G. Analysis of state-of-the-art smart metering communication standards. Proceedings of the 5th Young Researchers Symposium, 2010.

[CF+06] P.-A. Chirita, C.-S. Firan and W. Nejdl. Summarizing local context to personalize global web search. In Proceedings of the 15th ACM CIKM conference, pp. 287-296, 2006.

[CH+02] Cavanagh, S.M., Hanemann, W.M., and Stavins, R.N. Muffled Price Signals: Household Water Demand Under Increasing- Block Prices. FEEM Working Paper No. 40.2002, 2002.

[CK07] W. Chu and S.-S. Keerthi. Support Vector Ordinal Regression. Neural Computation, 19:792{815, 2007.

[CL+10] Campbell, T., Larson, E., Cohn, G., Froehlich, J., Alcaide, R., Patel, S. N. WATTR: a method for self-powered wireless sensing of water activity in the home. Proceedings of the 12th ACM international conference on Ubiquitous computing, pp. 169-172, 2010.

[CL+12] Chen, H., Lin, C., Hsieh, S.-H., Chao, H.-F-, Chen C.-S., Shiu R.-S., Ye, S.-R., Deng, Y.-C. Persuasive feedback model for inducint energy conservation behaviors of building users based on interaction with a virtual object. Energy and Buildings, 45, p.106-115, 2012.

[CP08] W. A. Chaovalitwongse and P. M. Pardalos. 2008. On the time series support vector machine using dynamic time warping kernel for brain activity classification. Cybernetics and Sys. Anal. 44, 1, 125-138, January 2008.

[Cr11] Crooke, J. (2011) Southern Water's Universal Metering Programme: 'intelligent' water meters employ Automatic Meter Reading (AMR) technology. Pg. 227 – 228. [Online], Available: http://www.waterprojectsonline.com/publication_date/2011.htm[8 June 2014].

[CR11] Carrico, A., Riemer, M. Motivating energy conservation in the workplace: An evaluation of the use of group-level feedback and peer education. Journal of Environmental Psychology, 31(1), p.1-13, 2011.

[CX+06] Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Huang and H.-W. Hon. Adapting ranking svm to document retrieval. In Proceedings of the 29th annual international ACM SIGIR conference, pp. 186-193, 2006.

[Da06] Darby, S. The effectiveness of feedback on energy consumption. Environmental Change Institute, University of Oxford, 2006.

[Da14] 1a. City of Dubuque (2014) Water – Treatment, Distribution and Customer Service. Online. Available at: http://www.cityofdubuque.org/494/Water [Access date: 26/06/2014].

[Db14] 1b. City of Dubuque (2014) Water Meter Replacement Project. Online. Available at: http://www.cityofdubuque.org/1304/Water-Meter-Replacement-Project [Access date: 26/06/2014].

[DC+08] J. Diez, J. J. del Coz, O. Luaces and A. Bahamonde. Clustering people according to their preference criteria. Expert Systems with Applications: An International Journal, v. 34, pp.1274-1284, 2008.

[Dc14] 1c. City of Dubuque (2014) Smarter Water. Online. Available at: http://www.cityofdubuque.org/1348/Smarter-Water/ [Access date: 26/06/2014].

[DD07] E. W. Dereszynski and T. G. Dietterich. Probabilistic Models for Anomaly Detection in Remote Sensor Data Streams. In Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence, 2007.

[De06] Delaware Public Service Commission. How can we manage what we cannot measure? Whitepaper on Advanced Metering. 2006.

[De08] Defra (2008) Future Water, the Government's water strategy for England. HM Government. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/69346/pb13562-future-water-080204.pdf.

[De11] Defra (2011). Water for Life. HM Government. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228861/8230.pdf.

[DF+03] Dalhuisen, J.M., Florax, R.J.G.M., de Groot, H.L.F., and Nijkamp, P. Price and income elasticities of residential water demand: a meta-analysis. Land Economics 79(2), p. 292-308, 2003.

[DG04] Jeffrey Dean and Sanjay Ghemawat. MapReduce: Simplified Data Processing on Large Clusters, OSDI, 2004.

[DL84] Daft, R.L., and Lengel, R.H. Information richness: a new approach to managerial behavior and organizational design. Research in organizational behavior (Homewood, IL: JAI Press), 6, p. 191–233, 1984.

[DN+97] Dandy, G., Nguyen, T., Davies, C., 1997. Estimating residential water demand in the presence of free allowances. Land Economics 73 (1), 125–139.

[DS+08] Z. Dou, R. Song, J.-R. Wen and X. Yuan. Evaluating the Effectiveness of Personalized Web Search. IEEE Transactions on Knowledge and Data Engineering, v. 21, pp. 1178-1190, 2008.

[DT+08] H. Ding and G. Trajcevski and P. Scheuermann and X. Wang and E. Keogh. Querying and mining of time series data: experimental comparison of representations and distance measures. Proc. VLDB

Endow. 1, 2, 1542-1552, August 2008

[Dw06] C. Dwork. Differential Privacy. In 33rd International Colloquium on Automata, Languages and Programming, part II, 2006.

[EB+04] Evans, R., Blotter, J., and Stephens, A. Flow Rate Measurements Using Flow-Induced Pipe Vibration. Journal of Fluids Engineering, 126(2), p. 280-285, 2004.

[EE+97] Espey, M., Espey, J., and Shaw, W.D. Price elasticity of residential demand for water: a meta-analysis. Water Resource Research, 33(6), p. 1369-1374, 1997.

[EG+10] Earleywine, M., Gonder, J., Markel, T., and Thornton, M. Simulated Fuel Economy and Performance of Advanced Hybrid Electric and Plug-in Hybrid Electric Vehicles Using In-Use Travel Profiles. 6th IEEE Vehicle Power and Propulsion Conference (VPPC), p. 1-6, 2010.

[EH+02] D. Eads and D. Hill and S. Davis and S. Perkins and J. Ma and R. Porterand and J. Theiler. Genetic Algorithms and Support Vector Machines for Time Series Classification. Proc. SPIE 4787. Pp. 74-85. July, 2002.

[EL+13] Erickson, T., Li, M., Kim, Y., Deshpande, A., Sahu, S., Chao, T., Sukaviriya, P., Naphade, M. The Dubuque Electricity Portal: Evaluation of a City-Scale Residential Electricity Consumption Feedback System, Conference on Human Interaction, p.1-10, 2013.

[EM05] EMWIS (2005) Local Water Supply, Sanitation and Sewage: Country Report Malta. Available at: http://www.emwis.org/documents/emwis-main-studies/studies-country/Malta_Final.pdf/download/1/Malta%20Final.pdf.

[Eu14] European Environment Agency (2014) Policies and measures to promote sustainable water use. Online. Available at: http://www.eea.europa.eu/themes/water/water-resources/policies-and-measures-to-promote-sustainable-water-use [Accessed: 04/07/2014]

[EW99] Evans, P., and Wurster, T. Getting real about virtual commerce. Harvard Business Review, 77(6), p. 84–94, 1999.

[FA+06] Fogarty, J., Au, C., and Hudson, S.E. Sensing from the Basement: A Feasibility Study of Unobtrusive and Low-Cost Home Activity Recognition. ACM Symposium on User Interface Software and Technology, p. 91-100, 2006.

[FA75] Fishbein, M., and Ajzen, I. Belief, Attitude, Intention, and Behavior! An Introduction to Theory and Research, Addison-Wesley Pub. Co., Reading Mass, 1975.

[FD+09] Froehlich, J., Dillahunt, T., Klasnja, P., Mankoff, J., Consolvo, S., Harrison, B., Landay, J. A. UbiGreen : Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits. Conference on Human Interaction, p.1043-1052, 2009.

[FF+11] Fadlullah, Z., Fouda, M., Kato, N., Takeuchi, A., Iwasaki, N., Nozaki, Y. Towards Intelligent Machine-to-Machine Communications in Smart Grid. IEEE Communications Magazine, 49(4), p.60-65, 2011

[FF+12] Froehlich, J., Findlater, L., Ostergren, M., Ramanathan, S., Peterson, J., Wragg, I., Larson, E., Fu, F., Bai, M., Patel, S., and Landay, J. The Design and Evaluation of Prototype Eco-Feedback Displays for Fixture-Level Water Usage Data. SIGCHI Conference on Human Factors in Computing Systems, p.

2367-2376, 2012.

[FF+12] Froehlich, J., Findlater, L., Ostergren, M., Ramanathan, S., Peterson, J., Wragg, I., Larson, E., Fu, F., Bai, M., Patel, S., and Landay, J. The Design and Evaluation of Prototype Eco-Feedback Displays for Fixture-Level Water Usage Data. SIGCHI Conference on Human Factors in Computing Systems, p. 2367-2376, 2012.

[FI03] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. The Journal of Machine Learning Research, v. 4, pp. 933-969, 2003.

[FL+09] Froehlich, J. E., Larson, E., Campbell, T., Haggerty, C., Fogarty, J., and Patel, S. N. HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. 11th international Conference on Ubiquitous Computing, p. 235-244, 2009.

[FS+09] Faruqui, A., Sergici, S., and Wood, L. Moving toward utility-scale deployment of dynamic pricing in mass markets. IEE Whitepaper, 2009.

[FS+13] Fielding, K., Spinks, A., Russell, S., McCrea, R., Sterwart, R., Gardner, J. An experimental test of voluntary strategies to promote urban water demand management. Journal of Environmental Management, 114, p.343-251, 2013.

[FS11] Frost & Sullivan(2011) Utilities push the smart water metering market in Europe finds. Available at: http://www.frost.com/prod/servlet/press-release.pag?docid=240769122 [access date: 04/07/14]

[Ga13] Gamification Wiki. What is Gamification? | Gamification.org. In Gamification Wiki | Gamification Wiki. Retrieved April 9, 2013, from http://gamification.org/wiki/Gamification.

[GB+11] G. Giannopoulos, U. Brefeld, T. Dalamagas, and T. Sellis. Learning to rank user intent. In Proceedings of the CIKM Conference, 2011.

[GE+83] Geller, E., Erickson, J., Buttram, B. Attempts to promote residential water conservation with educational, behavioral and engineering strategies. Population and Environment, 6(2), p.96-112, 1983.

[Ge01] P. Geurts. Pattern Extraction for Time Series Classification. In proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery, pp.115-127, 2001.

[GF+03] Gärling, T., Fujii, S., Gärling, A., & Jakobsson, C. (2003). Moderating effects of social value orientation on determinants of proenvironmental behavior intentions. Journal of Environmental Psychology, 23, 1–9.

[GG+12] Gölz, S., Götz, K., Klobasa, M., Schleich, J., Sunderer, G. (2012). Führt Verbrauchsfeedback zu Stromeinsparungen? Energiewirtschaftliche Tagesfragen. 62 (2012), 8, 32-33.

[GG09] Guille, C., Gross, G. A conceptual framework for the vehicle-to-grid (V3G) implementation. Energy policy, 37, 2009.

[GJ+11] Gamberini, L., Jacucci, G., Spagnolli, A., Corradi, N., Zamboni, L., Perotti, M, Cadenazzi, C., Mandressi, S., Tusa, G., Björkskog C., Salo, M, Aman, P. Saving is Fun: Designing a Persuasive Game for Power Conservation. Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, p. 16, 2011.

[GK+09] Grafton, R.Q:, Kompas, T., To, H., Ward, M., 2009, Residential water consumption: a cross country analysis. Environmental Economics Research Hub, research report no. 23, Canberra: Australian National University.

[Go10] Goldstein, H. (2010) Malta takes control. Spectrum, IEE 47 (6). 65 – 71.

[GR+10] Graving, J. S., Rakauskas, M. E., Manser, M. P., and Jenness, J. W. A Binary Response Method to Determine the Usability of Seven In-Vehicle Fuel Economy Displays. Human Factors and Ergonomics Society 54th Annual Meeting, p. 1546-1550, 2010.

[Gr13] Grohe. Grohe Ondus Digital product factsheet. Retrieved April 9, 2013, from http://www.grohe.com/us/5784/bathroom/bathroom-faucets/grohe-ondus-digital/

[GS+11] Gungor, V., Sahin, D., Kocak, T., Ergüt, S., Buccella, C., Cecati, C., Hancke, G. Smart grid technologies: Communication Technologies and Standards. IEEE transactions on Industrial informatics, 7(4), 529-539, 2011.

[GS99] V. Guralnik and J. Srivastava. Event Detection from Time Series Data. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 33-42, 1999.

[GW+10] Giurco, D., White, S. and Rodney, S. (2010). Smart Metering and Water End Use Data: Conservation and Privacy risks. Water. (2). 461 – 467.

[Ha07] Hassell, T. Promoting Behavioral Change in Household Water Consumption: Literature Review. URS, 2007.

[HB+01] Hunecke, M., Blöbaum, A., Matthies, E., & Höger, R. (2001). Responsibility and Environment: Ecological Norm Orientation and External Factors in the Domain of Travel Mode Choice Behaviour. Environment and Behaviour, 33, 830–852.

[HH07] Haustein, S., & Hunecke, M. (2007). Reduced use of environmentally friendly modes of transportation caused by perceived mobility necessities: An extension of the theory of planned behaviour. Journal of Applied Social Psychology, 37, 1856–1883.

[HK96] de Haan, G. & Kuckartz, U. (1996). Umweltbewusstsein. Denken und Handeln in Umweltkrisen. Opladen: Westdeutscher Verlag.

[HL12] Hall and Lobia, 2012 Hall, D and Lobina, E. (2012) Water companies and trends in Europe 2012. Public Services International Research Unit.http://www.epsu.org/IMG/pdf/2012_Water_companies-EWCS.pdf

[HM+86] Hutton, R.B., Mauser, G.A., Filiatrault, P., and Ahtola, O.T. Effects of Cost-Related Feedback on Consumer Knowledge and Consumption Behavior: A Field Experimental Approach. Journal of Consumer Research, 13(3), p. 327-336, 1986.

[HM10] Ham, J., Midden, C. A Persuasive Robotic Agent to Save Energy: The Influence of Social Feedback, Feedback Valence and Task Similarity on Energy Conservation Behavior. Social Robotics, p.335-344, 2010.

[HM98] Homburg, A. & Matthies, E. (1998). Umweltpsychologie: Umweltkrise, Gesellschaft und Individuum. Weinheim und München: Juventa Verlag.

[HN91] Hopper, J. R. & Nielsen, J. M. (1991). Recycling as altruistic behaviour. Normative and behavioural strategies to expand participation in a community recycling program. Environment and Behaviour, 23, 195–220.

[Ho07] Holmes, T. Eco-visualization: Combining Art and Technology to Reduce Energy Consumption. Proceedings of the 6th ACM SIGCHI Conference on Creativity&Cognition, p.152-162, 2007.

[HR89] van Houwelingen, H., and van Raaij, W. The Effect of Goal-Setting and Daily Electronic Feedback on In-Home Energy Use. Journal of Consumer Research, 16(1), p. 98-105, 1989.

[HW+08] Herber, W., Wagner, H., Roth, U. (2008): Die Wasserbedarfsprognose als Grundlage für den Regionalen Wasserbedarfsnachweis der Hessenwasser GmbH & Co. KG, Wasser Abwasser 149 (5), 2008

[IB11] IBM Research (2011) Smart Water pilot study report: City of Dubuque, USA. IBM: Available at: http://www.cityofdubuque.org/DocumentCenter/Home/View/3116.

[IJ06] Inman D., and Jeffrey P. A review of residential water conservation tool performance and influences on implementation effectiveness. Urban Water Journal, 3(3), p. 127 - 143, 2006.

[IM+10] Ijab, M.T., Molla, A., Kassahun, A.E., and Teoh, S.Y. Seeking the 'Green' in 'Green IS': A Spirit, Practice and Impact Perspective, Pacific-Asia Conference of Information Systems, 2010.

[Jo02] T. Joachims. Optimizing search engines using clickthrough data. In Proceedings of the ACM SIGKDD conference, pp. 133-142, 2002.

[Jo06] T. Joachims. Training Linear SVMs in Linear Time. In Proceedings of ACM SIGKDD conference, 2006.

[JW+11] Jenkin, T.A., Webster, J., and McShane, L. An Agenda for 'Green' Information Technology and Systems Research, Information and Organization, 21(1), p. 17-40, 2011.

[Ka96] Kals, E. (1996). Verantwortliches Umweltverhalten. Weinheim: Psychologie Verlags Union.

[KC09] J.-W. Kim and K.-S. Candan. Skip-and-prune: cosine-based top-k query processing for efficient context-sensitive document retrieval. In Proceedings of the 35th SIGMOD conference, pp. 115-126, 2009.

[KD+05] Kurz, T., Donague, N., Walker, I. Utilizing a Social-Ecological Framework to Promote Water and Energy Conservation: A Field Experiment. Journal of Applied Social Psychology, 35(6), p.1281-1300, 2005.

[KG+08] Kenney, D.S., Goemans, C., Klein, R., Lowrey, J. and Reidy, K. Residential water demand management: Lessons from Aurora, Colorado. *Journal of the American Water Resources Association,* 44(1), p. 192-207, 2008.

[KG03] Kaiser, F. G., & Gutscher, H. (2003). The proposition of a general version of the theory of planned behaviour: Predicting ecological behaviour. Journal of Applied Social Psychology, 33, 586–603.

[KG09] Kappel, K., and Grechenig, T. "show-me": Water Consumption at a Glance to Promote Water Conservation in the Shower. 4th International Conference on Persuasive Technology, p. 26:1-26:6, 2009.

[KK+04] Kenney, D.S., Klein, R.A., and Clark, M.P. Use and Effectiveness of Municipal Water Restrictions

During Drought in Colorado. Journal of the American Water Resources Association, 40(1), p. 77-87, 2004.

[KK+07] Klein, B., Kenney, D., Lowrey, J., and Goemans, C. Factors influencing residential water demand: a review of the literature. Working paper of the University of Colorado, 2007.

[KL02] E. Keogh, S. Lonardi, and B. Chiu. Finding Surprising Patterns in a Time Series Database in Linear Time and Space. In proceedings of the 8th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining, pp. 550-556, July, 2002.

[KL06] E. Keogh and J. Lin and A. Fu. Finding the Most Unusual Time Series Subsequence: Algorithms and Applications. Knowledge and Information Systems (KAIS), Springer-Verlag, 2006

[KM+09] Argyro Kampouraki and George Manis and Christophoros Nikou. 2009. Heartbeat time series classification with support vector machines. Trans. Info. Tech. Biomed. 13, 4, 512-518, July 2009.

[KP+99] Kalafatis, S. P., Pollard, M., East, R. & Tsogas, M. H. (1999). Green marketing and Ajzen's theory of planned behaviour: a cross-market examination. Journal of Consumer Marketing, 16, 441–460.

[KP10] Kuznetsov, S., and Paulos, E. UpStream: Motivating Water Conservation with Low-Cost Water Flow Sensing and Persuasive Displays. SIGCHI Conference on Human Factors in Computing Systems, p. 1851–1860, 2010.

[KS+08] Kim, Y., Schmid, T., Charbiwala, Z.M., Friedman, J. and Srivastava, M.B. NAWMS: Non-Intrusive Autonomous Water Monitoring System. ACM Conference on Embedded Network Sensor Systems, p. 309-322, 2008.

[LC06] Lam, S. P., & Chen, J. K. (2006). What makes costumers bring their bags or buy bags from the shop? A survey of customers at a Taiwan hypermarket. Environment and Behavior, 38, 318–322.

[LD+05] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan. Incognito: Efficient Full-domain k-Anonymity, in SIGMOD, pp. 49–60, 2005.

[LD+06] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan. Mondrian Multidimensional k-Anonymity, in ICDE, 2006.

[LE+08] Lee, Y-W., Eun, S., Oh, S-H. Wireless Digital Water Meter with Low Power Consumption for Automatic Meter Reading. Proceedings of the Internation Conference on Convergence and Hybrid Information Technology, p.639-645 2008.

[LF+12] Larson, E., Froehlich, J., Campbell, T., Haggerty, C., Atlas, L., Fogarty, J., and Patel, S. N., Disaggregated water sensing from a single, pressure-based sensor: An extended analysis of HydroSense using staged experiments. Pervasive and Mobile Computing, 8(1), p. 82-102, 2012.

[LH+11] Laschke, M., Hassenzahl, M., Diefenbach, S., and Tippkämper, M. With a little help from a friend: A Shower Calendar to save water. CHI '11 Extended Abstracts on Human Factors in Computing Systems, p. 633-646, 2011.

[Li13] Liu, Q. BLuePot: An Ambient Persuasive Approach to Domestic Energy Saving. Proceedings of the 4th IEEE International Conferenc on Consumer Electronics (ICCE), p.106-107, 2013.

[LL+12] T. Li, N. Li, J. Zhang and I. Molloy. Slicing: A New Approach for Privacy Preserving Data Publishing, in TKDE, (vol. 24 no. 3) pp. 561-574, 2012.

[LM+83] Lutz, R. J., Mackenzie, S. B., and Belch, G. E. Attitude Toward the Ad as a Mediator of Advertising Effectiveness: Determinants and Consequences, in Advances in Consumer Research (10), pp. 532–539, 1983.

[LO+08] J. Li, B. C. Ooi, and W. Wang. Anonymizing streaming data for privacy protection. In ICDE, pages 1367-1369, 2008.

[LS+11] Loock, C., Staake, T., Landwehr, J. Green IS Design and Energy Conservation: An Empirical Investigation of Social Normative Feedback. Thirty Second International Conference on Information Systems, 2011.

[LS+13] Loock, C., Staake, T., and Thiesse, F. Motivating Energy-Efficient Behavior with Green IS: An Investigation of Goal Setting and the Role of Defaults. MIS Quarterly, 37(4), 1313-1332., 2013.

[Lu93] Luitzenhiser, L. Social and Behavioral Aspects of Energy Use. Annual Review Energy Environment, 18, p-247-89, 1993.

[LV04] J. Lin and M. Vlachos and E. Keogh and D. Gunopulos. Iterative Incremental Clustering of Time Series. IX Conference on Extending Database Technology (EDBT), March, 2004.

[LW+07] X. Li, N. Wang and S.-Y. Li. A fast training algorithm for svm via clustering technique and Gabriel graph. In Proceedings of the International Conference on Intelligent Computing, 2007.

[Ly13] Lyonnaise des Eaux (2013) Press Release: A Lyonnaise des Eaux, A Suez Environment subsidiary wins a remote meter reading contract in Mulhouse and goes over more than the one million mark for smart meters installed in Europe. Available at: http://www.suez-environnement.fr/wp-content/uploads/2013/10/CP-Mulhouse_VA.pdf

[Ma05] Matthies, E. (2005). Wie können PsychologInnen ihr Wissen besser an die PraktikerIn bringen? Vorschlag eines neuen integrativen Einflussschemas umweltbewussten Alltagshandelns. Umweltpsychologie, 9, 62–81.

[Ma07] Martinez-Espineira, R., 2007, An Estimation of Residential Water using co-integration and error correction techniques. Journal of Applied Economics 10 (1): 161–184

[MA07] Messner, F., Ansmann, T., 2007, Wassernutzung der privaten Haushalte in Leipzig – Einflussfaktoren der Wassernachfrage und Bedeutung der individuellen Wahrnehmung dieser Faktoren durch die Wassernutzer. UFZ-Bericht 01/2007, Helmholtz-Zentrum für Umweltforschung UFZ, Leipzig.

[Ma10] Marchment Hill Consulting (2010) Smart Water Metering Cost Benefit Study. Victorian Water Trust. Available at: http://www.swan-forum.com/uploads/5/7/4/3/5743901/smart_metering_cost_benefit.pdf.

[MC+13] Mitchell, D., Cubed, M., Chesnutt, T. Evaluation of East Bay Municipal Utility District's Pilot of WaterSmart Home Water Reports. Policy analysis fort he public and private sectors, 2013.

[MC79] McClelland, L. and Cook, S.W. Energy conservation effects of continuous in-home feedback in all-electrichomes. Journal of Environmental Systems, 9, p. 169-73, 1979-80.

[MC91] Maddock, R., Castaño, E., 1991. The welfare impact of rising block pricing: electricity in Colombia. Energy Journal 12 (4), 65–77

[MD+04] Mayer, P. W., DeOreo, W. B., Towler, E., Martien, L., Lewis, D. M. The Impacts of High Efficiency Plumbing Fixture Retrofits in Single-Family Homes. Tampa Water Department and The U.S. Environmental Protection Agency, 2004.

[MD+10] Ehrhardt Martinez, K., Donnelly, K. A., and Laitner, S. Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities. ACEEE Research Report E105, 2010.

[MD+99] Mayer, P.W., DeOreo, W.B., Kiefer, J. Opitz, E., Dziegielewski, B., and Nelson, J.O. Residential End Uses of Water. American Water Works Association, study, 1999.

[Me10] Melville, N. P. Information Systems Innovation for Environmental Sustainability, Management Information Systems Quarterly, 34(1), p. 1-21, 2010

[MG+06] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. L-Diversity: Privacy Beyond k-Anonymity. In ICDE, 2006.

[MM+02] P. Melville and R. Mooney and R. Nagarajan. Content-boosted collaborative filtering for improved recommendations. In 18th National Conference on Artificial Intelligence (pp. 187-192), 2002.

[MN04] Martinez-Espineira, R., Nauges, C., 2004. Is all domestic water consumption sensitive to price control? Applied Economics 36: 1697–1703

[Mo07] B. Mobasher. Data mining for web personalization. In The Adaptive Web, 90-135, 2007.

[Mo09] Molla, A. The Reach And Richness Of Green IT: A Principal Component Analysis. Australian Conference on Information Systems, p. 754-764, 2009.

[Mo87] Moncur, J.,1987, Urban water pricing and drought management. Water Resources Research 23 (3): 393–398.

[MR99] R. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In ACM SIGIR'99, Workshop on Recommender Systems: Algorithms and Evaluation, 1999.

[MS02] Micallef, P. and Schembri, R. (2002) Water evaluation and cost recovery mechanisms in the developing countries of the Mediterranean region. Malta case. In : Hamdy A. (ed.), Lacirignola C. (ed.), Lamaddalena N. (ed.). Water valuation and cost recovery mechanisms in the developing countries of the Mediterranean region. Bari : CIHEAM, 2002. p. 149-158.

[MS99] McKenzie-Mohr, D., Smith, W. Fostering sustainable behavior: An introduction to community-based social marketing (2nd ed.). Gabriola Island, British Columbia, Canada, 1999.

[MW+09] Meschtscherjakov, A., Wilfinger, D., Scherndl, T., and Tscheligi, M. Acceptance of Future Persuasive In-Car Interfaces Towards a More Economic Driving Behaviour. First International Conference on Automotive User Interfaces and Interactive Vehicular Applications, p. 81-88, 2009.

[NC+07] M. Nergiz, C. Clifton, and A. Nergiz, Multirelational k-anonymity, in ICDE, pp. 1417 – 1421, 2007.

[NC07] M. E. Nergiz and C. Clifton. Thoughts on k-anonymization, Data and Knowledge Engineering, vol. 63, no. 3, pp. 622–645, 2007.

[Ni92] Nieswiadomy, M. Estimating urban residential water demand: effects of price structure, conservation, and education. Water Resources Research, 28(3), p.609-615, 1992.

[NL+11] M.Naphade and D. Lyons and C.A. Kohlmannand C. Steinhauser. Smart Water Pilot Study Report, IBM Research - Watson, 2011.

[NL+11] Naphade, M., Lyons, D., Kohlmann, C.A., and Steinhauser C. Smart Water Pilot Study Report, IBM Research - Watson, 2011.

[NM89] Nieswiadomy, M. L., and Molina, D. J. Comparing Residential Water Estimates Under Decreasing and Increasing Block Rates Using Household Data. Land Economics, 65(3), p. 280–289, 1989.

[No76] Nordin, J.A., 1976. A proposed modification on Taylor's demand–supply analysis: comment. The Bell Journal of Economics 7 (2), 719–721

[NR+10] Neunteufel, R., Richard, L., Perfler, R., 2010, Studie Wasserverbrauch und Wasserbedarf. Teil 1: Literaturstudie zum Wasserverbrauch – Einflussfaktoren, Entwicklung, Prognosen. Lebensministerium, Wien.

[NT00] Nauges, C., Thomas, A., 2000, Privately Operated Water Utilities, Municipal Price Negotiation, and Estimation of Residential Water Demand: The Case of France. Land Economics 76 (1): 68–85.

[NT03] Nauges, C., Thomas, A., 2003, Long-run Study of Residential Water Consumption. Environmental and Resource Economics 26: 25–43.

[OE08] Household Behavior and the Environment, Reviewing the Evidence. OECD 2008, p. 264, 2008.

[Of12] Ofwat (2012) Key Milestones Ofwat and the water and sewerage sectors. Available at: http://www.ofwat.gov.uk/mediacentre/fastfacts/prs_web_timeline.pdf [access date: 04/07/2014]

[OH09] Oinas-Kukkonen, H., and Harjumaa, M. Persuasive Systems Design: Key Issues, Process Model, and System Features. Communications of the Association for Information Systems, 24(28), 2009.

[Os97] Osterhus, T.L. Pro-social consumer influence strategies: When and how do they work? Journal of Marketing, 61(4), p. 16-29, 1997.

[Pa14] Pace, R. (2014) The AMM Project in Malta. Presentation: Water Loss 2014, Vienna.

[PK+10] Parikh, P. P., Kanabar, M. G., Sidhu, T. S. Opportunities and challenges of wireless communication technologies for smart grid applications. IEEE Power and Energy Society General Meeting, pp. 1-7, 2010.

[PK+11] Petkov, P., Krcmar, H., Foth, M., and Medland, R. Engaging Energy Saving through Motivation-Specific Social Comparison. CHI '11 Extended Abstracts on Human Factors in Computing Systems, p. 1945-1950, 2011.

[PP+09] D. Preston and P. Protopapas and C. Brodley. Event Discovery in Time Series. In SIAM SDM09, 2009.

[PS+03] Poortinga, W., Steg, L., Vlek, C., and Wiersma, G. Household preferences for energy-saving measures: A conjoint analysis. Journal of Economic Psychology, 24(1), p. 49-64, 2003.

[PS+07] Petersen, J., Shunturov, V., Janda K., Platt, G., Weinberger, K. Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives. International Journal of Sustainability in Higher Education, 8(1), p.16-33, 2007.

[QZ+07] T. Qin, X.-D. Zhang, D.-S. Wang, T.-Y. Liu, W. Lai and H. Li. Ranking with multiple hyperplanes. In Proceedings of the 30th annual international ACM SIGIR, pp. 279-286, 2007.

[Ra12] RAND Europe. What Works in Changing Energy – Using Behaviors in the Home? A Rapid Evidence Assessment. Department of Energy and Climate Change, 2012.

[Re06] Reynaud A. 2006. Social Policy, Regulation and the Private Sector in Water Supply: FranceCase Study. Final report for the United Nations Research Institute for Social Development. Geneva. 104 pages.

[RG+06] P.P. Rodrigues, J. Gama, and J.P. Pedroso. ODAC: Hierarchical Clustering of Time Series Data Streams. Proc. Sixth SIAM Int'l Conf. Data Mining, pp. 499-503, Apr. 2006.

[RG00] Renwick, M.E. and R.D. Green, 2000. Do Residential Water Demand Side Management Policies Measure Up? An Analysis of Eight California Water Agencies. Journal of Environmental Economics and Management 40:37-55.

[Sa14] 1a.Suez Environment (2014) Projects and key figures. Online. Available at: http://www.suez-environnement.com/water/projects-key-figures/. Access date: 09/07/2014]

[SB04] Satterfield, Z., Bhardwaj, V. Water Meters. Tech Brief, 4(2), 2004.

[Sc10] Schleich, J., 2010. How low can you go? Exploring threshold levels for residential water demand in Germany. ???

[Sc77] Schwartz, S. H. (1977). Normative influences on altruism. Advances in Experimental Social Science, 10, 221–279.

[SD00] Saleth, R. M., Dinar, A., 2000, Urban thirst: water supply augmentation and pricing policy in Hyderabad City, India. World Bank, Washington, DC.

[SD85] Schefter, J.E., David, E.L., 1985. Estimating residential water demand under multi-tariffs using aggregate data. Land Economics 61 (3), 272–280

[SF11] Staake, T., Fischli, C., Stiefmeier, T., Tröster, G., and Fleisch, E. Smart Water Metering: Consumption Feedback at the Tap. Energy - Economy - Society Research Programme (EES), Swiss Federal Office of Energy, 2011.

[SG+11] S. Spiegel, J. Gaebler, A. Lommatzsch and E. De Luca and S. Albayrak. Pattern Recognition and Classification for Multivariate Time Series. In Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data, pp. 34-42, 2011.

[SH+04] K. Sugiyama, K. Hatano and M. Yoshikawa. Adaptive web search based on user profile constructed without any effort from users. In Proceedings of the 13th international conference on World Wide Web, pp. 675-684, 2004.

[SH09] Schleich, J., Hillenbrand, T., 2009, Determinants of Residential Water Demand in Germany. Ecological Economics 68: 1756–1769

[SH82] Schwartz, S. H., & Howard, J. A. (1982). Helping and cooperation: A self-based motivational model. In: V. J. Derlega,& J. Grzelak (eds.), Cooperation and helping behavior: Theories and research. New York: Academic press, 327–353.

[Si55] Simon, H. A behavioral model of rational choice. The Quarterly Journal of Economics, 69(1), p-99-118, 1955.

[SK+12] Schleich, J., Klobasa, M., Gölz, S., Brunner, M. (2012). Effects of Feedback on Residential Electricity Demand - Findings from a Field Trial in Austria. Working Paper Sustainability and Innovation No. S 8/2012. Karlsruhe : Fraunhofer ISI

[SK13] Schleich, J., Klobasa, M. (2013). How Much Shift in Demand? Findings from a Field Experiment in Germany. eceee 2013 Summer Study on Energy Efficiency. Proceedings : 3-8 June 2013, Club Belambra Les Criques, Presqu,'île de Giens, Toulon/Hyères, France. Stockholm : eceee, 1919-1925.

[Sm06] Smets, H. (2006) Implementing the right to water in France. Paper prepared for workshop entitled 'Legal aspects of Water Sector Reforms'. Available at: http://www.ielrc.org/activities/workshop_0704/content/d0723.pdf.

[SN+00] Syme, G.J., Nancarrow, B.E., and Seligman, C. The Evaluation of Information Campaigns to Promote Voluntary Household Water Conservation. Evaluation Review, 24(6), p. 539-578, 2000.

[SN+07] Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. The constructive, destructive, and reconstructive power of social norms. Psychological Science, 18(5), p. 429-434, 2007.

[ST+05] X. Shen, B. Tan, and C. Zhai. Context-sensitive information retrieval using implicit feedback. In Proceedings of the 28th annual international ACM SIGIR conference, pp. 43-50, 2005.

[St08] Steg, L. Promoting household energy conservation. Energy Policy, 36(12), p. 4449-4453, 2008.

[St10] Staddon, C. (2010) Do Water Meters Reduce Domestic Consumption? A Summary of Available Literature. Defra, London, UK.

[St11] Strengers, Y. Negotiating everyday life: The role of energy and water consumption feedback. Journal of Consumer Culture, 11(3), p.319-338, 2011.

[SV09] Steg, L., and Vlek, C. Encouraging pro-environmental behaviour: An integrative review and research agenda. Journal of Environmental Psychology, 29(3), p. 308-317, 2009.

[SW+09] Stewart, R. A., Willis, R., Capati, B. Closing the loop on water planning: an integrated smart metering and web-based knowledge management system approach. ICA 2009. International Water Association, 2009.

[SW+09] Shiraishi, M., Washio, Y., Nakajima, T. Using Individual, Social and Economic Persuasion Techniques to Reduce CO Emissions in a Family Setting, Proceedings of Persuasive, 2009.

[SW+10] Stewart, R. A., Willis, R., Giurco, D., Panuwatwanich, K., & Capati, G. Web-based knowledge management system: linking smart metering to the future of urban water planning. Australian Planner, 47(2), 66-74, 2010.

[SW+13] Stewart, R. A., Willis, R. M., Panuwatwanich, K., Sahin, O. Showering behavioural response to alarming visual display monitors: longitudinal mixed method study. Behaviour & Information

Technology, 32(7): 695–711, 2013.

[Sw02] L. Sweeney. K-Anonymity: A Model for Protecting Privacy. In IJUFKS, vol. 10, no. 5, 2002.

[SW09] Southern Water (2009) Final Business Plan. Available online: http://www.southernwater.co.uk/pdf/about-us/publications/business-plan/FBPSectionCPub_sec4.pdf.

[Sw12] Swan, M. Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. Journal of Sensor and Actuator Networks, 1(3), pp. 217–253, 2012.

[SW13] R. Silipo and P. Winters. Big Data, Smart Energy, and Predictive Analytics. Time Series Prediction of Smart Energy Data. KNIME.com, white paper, 2013.

[SW14] Southern Water (2014) About us. Online. Available at: http://www.southernwater.co.uk/about-us/about-southern-water/ [Access date:29/06/2014].

[Ta75] Taylor, L.D., 1975. The demand for electricity: a survey. The Bell Journal of Economics 6 (1), 74–110

[TD+08] J. Teevan, S.-T. Dumais and D.-J. Liebling. To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent. In Proceedings of the 31th annual international ACM SIGIR conference, pp. 163-170, 2008.

[Th99] Thøgersen, J. (1999). The ethical consumer. Moral norms and packaging choice. Journal of Consumer Policy, 22, 439–460.

[TP+04] Tonglet, M., Phillips, P. S., & Read, A. D. (2004). Using the theory of planned behaviour to investigate the determinants of recycling behaviour: A case study from Brixworth, UK. Resources, Conservation and Recycling, 41, 191–214.

[TT+13] Tiefenbeck, V., Tasic, V., and Schöb, S. Mechatronics to drive environmental sustainability!: measuring, visualizing and transforming consumer patterns on a large scale, in Energy Informatics, Vienna, Austria, pp. 1–6, 2013.

[TW+05] Turner, A., White, S., Beatty, K., and Gregory, A. Results of the largest residential demand management program in Australia. Sydney: Institute for Sustainable Futures, University of Technology, 2005.

[TW14] Thames Water (2014) It's important to save water. Available at: http://www.thameswater.co.uk/metering/17011.htm. [Access date:29/06/2014]

[UW14] Utility Week (2014) A smart move for water. Online. Available at: http://www.utilityweek.co.uk/news/a-smart-move-for-water/975152#.U7--IPldXA4 [access date: 11/07/2014]

[VM+13] Vinod Kumar Vavilapalli, Arun C. Murthy, Chris Douglas, et al. Apache Hadoop YARN: Yet Another Resource Negotiator, SOCC, 2013.

[VV+13] Van der Sanden, M. C. A., Van Dam, K. H., Stragier, J., and Kobus, C. B. A. Simulation Based Decision Support for Strategic Communication and Marketing Management Concerning the Consumer Introduction of Smart Energy Meters, in Journal of Communication Studies (6:1), pp. 75–104, 2013.

[Wa12] Waterwise (2012) Desalination is not the answer. Online. Available at: http://www.waterwise.org.uk/news.php/29/desalination-is-not-the-answer. [Access date:]

[Wa13] Waterwise (2013) Water Efficiency and Drought Communications report. Waterwise.Available at: http://www.waterwise.org.uk/data/2013_Waterwise_Drought_Report.pdf

[WD14] The Water debate (2014) 'Smart Water' technological intelligence applied to water management. Governance and Politics. Available from: http://www.eau-debat.fr/2014/06/18/23185-smart-water-l-intelligence-technologique-appliquee-a-la-gestion-de-l-eau [access date: 10/07/2014].

[We08] Wetherall, B. Final report of the Eco-Pioneer Pilot Program. 2008

[We13] Welsh Water. Water Facts. Retrieved April 11, 2013 from http://www.dwrcymru.co.uk/welsh/waterefficiency/school/_pdf/waterfacts.pdf

[WK06] L. Wei and E. Keogh. Semi-supervised time series classification. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 748-753, 2006.

[WL+12] Weiss, M., Loock, C., Staake, T., Mattern, F., Fleisch, E. Evaluating Mobile Phones as Energy Consumption Feedback Devices. Mobile and Ubiquitous Systems: Computing, Networking, and Services, p.63-77, 2012.

[WS+10] Willis, R.M., Stewart, M.R.A., Panuwatwanich, K., Jones S., and Kyriakides, A. Alarming visual display monitors affecting shower end use water and energy conservation in Australian residential households. Resources, Conservation and Recycling, 54(12), p. 1117-1127, 2010.

[WS+11] Willis, R.M., Stewart, R.A., Panuwatwanich, K., Williams, P.R. and Hollingsworth A.L. Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. Journal of Environmental Management, 92(8), p. 1996-2009, 2011.

[WS+12] M. Way and J. Scargle and K. Ali and A. Srivastava. Advances in Machine Learning and Data Mining for Astronomy. Chapman and Hall/CR, March, 2012

[WS04] WSC (2004) Annual Report 2004 – Operations. Online. http://www.wsc.com/mt/annual_reports/2004/part1.pdf.

[Ws13] WaterSense, U.S. Environmental Protection Agency. (April 8, 2013). US Indoor Water Use | WaterSense | US EPA. In US Environmental Protection Agency. Retrieved April 9, 2013, from http://www.epa.gov/WaterSense/pubs/indoor.html.

[WS13] Woods, E. and Strother, N. (2013) Executive Summary: Smart Water Networks Smart Water meter, communication infrastructure, network monitoring and automation technologies, and data management and analytics: Market Analysis and forecast. Navigant Research. Available at: http://www.navigantresearch.com/wp-assets/uploads/2013/12/SWAT-13-Executive-Summary.pdf.

[XT06] X. Xiao and Y. Tao. Anatomy: simple and effective privacy preservation, in VLDB, pp. 139–150, 2006.

[XT07] X., Xiao and Y. Tao. m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets, in SIGMOD, pages 689-700, 2007.

[YA83] Yates, S., Aronson, E. A social psychological perspective on energy conservation in residential buildings. American Psychologist, 38(4), p.435-333, 1983.

[ZC+07] Z. Zheng, K. Chen, G. Sun and H. Zha. A regression framework for learning ranking functions using relative relevance judgments. In Proceedings of the 30th annual international ACM SIGIR conference, pp.287-294, 2007.

[ZH+09] B. Zhou, Y. Han, J. Pei, B. Jiang, Y. Tao and Y. Jia. Continuous Privacy Preserving Publishing of Data Streams. In EDBT, pages 648-659, 2009.

[Zy11] Van Zyl, J. Introduction to Integrated Water Meter Management. Water Research Commission, Gezine, South Africa, 2011.