Project no. 732027

VIRT-EU

Values and ethics in Innovation for Responsible Technology in EUrope

Horizon 2020

ICT-35-2016
Enabling responsible ICT-related research and innovation
Start date: 1 January 2017 – Duration: 36 months

D2.1
Blog posts and multi-media material summarizing preliminary empirical and policy findings

Due date: 30 September 2017

Re-submission date: 28 September 2018

Number of pages: 31

Lead beneficiary: CIID

Author(s): All partners

Project Consortium

Beneficiary no.	Beneficiary name	Short name
1 (Coordinator)	IT University of Copenhagen	ITU
2	London School of Economics	LSE
3	Uppsala Universitet	UU
4	Politecnico Di Torino	POLITO
5	Copenhagen Institute of Interaction Design	CIID
6	Open Rights Group	ORG

Dissemination Level

PU	Public	**X**
CO	Confidential, only for members of the consortium (including the Commission Services)	
EU-RES	Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)	
EU-CON	Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)	
EU-SEC	Classified Information: SECRET UE (Commission Decision 2005/444/EC)	

Dissemination Type

R	Document, report	
DEM	Demonstrator, pilot, prototype	
DEC	Websites, patent filling, videos, etc.	**X**
O	Other	
ETHICS	Ethics requirement	

Table of Contents

Executive summary

Deliverable 2.1 “Blog posts and multi-media material summarising preliminary empirical and policy findings for developer communities under study and other interested stakeholders, disseminated through a variety of social media channels” sits in the Research WP 2 “Domain Analysis” and in particular inside Task 2.6 aiming at a synthesis of findings and a formulation of domain requirements.

Given the progress of Work Package 2, CIID, together with the other VIRT-EU partners, has gathered a series of blog posts that summarise the main findings in the project thus far - specifically in terms of the legal and ethnographic studies. Furthermore, in this report we share preliminary methods for visualising the learnings of the legal and ethnographic teams - through connection-based value maps and roadmap-like diagrams. These initial visualisations are foundations for continued interactive development of artefacts to share how different values and concepts are the basis for our project. The visual representations of abstract values and ethical approaches allow our partnership to identify gaps and overlaps, as well as to possibly form new understandings of how to practically implement tools for ethical reflection and self-assessment during the design process. The diagrams and mappings will continue to evolve as we continue to improve how we communicate our project both to the broader public and to our internal collaborators - whether the groups who join us for co-design workshops or the stakeholders who are part of our stakeholder workshops. By placing the simplified depictions on paper, we can better pull them apart and remake them - a core value within our overall VIRT-EU project and our focus on co-creation.

The results presented in this report will be used for dissemination to the IoT developers communities and to the wide public via blogs and other types of social media channels, as well as first low-fidelity, paper-based prototypes to facilitate the partnership dialogue and continuation co-design workshops with IoT developers (WP3).

SECTION 1: Blog Posts

As described in the executive summary, one of our primary goals is to ensure that the complexity of academic and theoretical work conducted by the VIRT-EU partners in the first year of the projects is communicated as broadly as possible, including to non-academic audiences. As part of this effort the NEXA center at POLITO, ITU and CIID iteratively developed a series of blogposts to present the work done in Work Package 2. The work of converting a 120 page deliverable to short popular appeal texts is not trivial and required expertise in interaction design and media communication as well as familiarity with the non-academic audiences relevant to the project. The three blogposts are part of a series of discussions emerging from D2.2 and oriented towards IoT practitioners, civil society actors and other interested parties. These are presented here as examples of the text already posted and in the process of being posted to the VIRT-EU website (please note that due to current website migration, the content on the website is in flux).

BLOG POST 1
Searching for moral reasoning in the IoT
An exploratory analysis of IoT communities and Manifestos

Torey Rubin is one of the thousand customers left in the dark by Emberlight, a start-up that produced a smart socket which used to transform a traditional bulb into a smart light. On the 16th of November 2017, the start-up emailed its customers notifying that it was going out of business because of the pressure from bigger competitors (figure 1).

Dear Emberlight Customers and Partners,

Thank you for joining us in our journey to build simple smart lighting products. In 2013, we saw that home automation was a confusing and complicated experience, and we set out to create a simple product that was easy to install and use.

In 2014, we launched the Emberlight Socket on Kickstarter with a humble goal of $50,000. We were blown away when we reached over $300,000 in preorder sales during the 6 week campaign. We were fortunate to raise a seed round from amazing investors to bring the product to market, and those investors stood behind us through thick and thin. We knew that our product vision was mainstream when made it on the shelves of 270 Target stores across the country.

Since our Kickstarter campaign, we have seen at least 4 identical products hit the market, along with hundreds of other smart lighting products. For customers, this is great as it has driven down prices and increased choice. However, it has proven to be a challenging environment for a small startup. The reality is that we need a lot more capital to survive in this market segment against heavy hitters like Philips, GE and others. Unfortunately, we have not been able to raise the required financing to continue our journey.

With that, I'm sad and sorry to say that emberlight is shutting down. We are deeply grateful to the customers, fans, investors and partners that believed in emberlight's mission and helped us get as far as we did.

Questions that you may have:

Will you continue to sell emberlight Sockets? We are no longer selling emberlight Sockets. We have also informed retailers that are currently carrying emberlight.

Will my emberlight Socket continue to work? Since emberlight is an "Internet of Things" product, the device works through 3rd party servers that are hosted in the cloud which have a monthly fee. We do have some credits remaining with our server provider which should enable your devices to work for 3-4 more months, save any major bugs.

Can I return my emberlight Sockets? Emberlight cannot accept returns at this time. If you purchased your product through a retailer, you will need to contact them about their return policy.

Will you continue to offer technical support? Unfortunately, we no will no longer be able to resolve technical bugs related to the product. We will keep our support page up for 3-4 more months.

Figure 1 Emberlight announcement

In the communication sent in a FAQ-like-style, the question that was taking in hostage Torey's mind (Will my Emberlight Socket continue to work?) got the following answer:

"Since Emberlight is an "Internet of Things" product, the device works through 3rd party servers that are hosted in the cloud which have a monthly fee. We do have some credits remaining with our server provider which should enable your devices to work for 3-4 more months, save any major bugs."

Routing information via cloud is a common feature of IoT-related technologies. The cloud affords small businesses and start-ups as Emberlight with straightforward advantages, such as saving money for those resources-consuming operations as data storage and servers security updates. However, it also comes with some downsides, as experienced by Torey. In light of this, might there have been alternative solutions that could have prevented Torey from being left in the dark? Why didn't Emberlight designers opt for the bluetooth protocol that would have made the socket functioning local thus not relying solely on a third-party server? Were designers aware of the potential **trade-offs** they were making while creating the connected device?
Answering these questions required us to map the kind of **informal moral reasoning** designers undertake when developing IoT products and services.

Why we want to map "moral reasoning" in the IoT domain

When Torey took part to the first round of pre-sales orders Emberlight launched on Kickstarter, he gave little or no account to **the design features** of the socket. He was probably keen about the benefits the smart socket may have guaranteed, either by keeping the bulb separated from the smarts (a beneficial choice in case the bulb breaks or burns out) or by allowing him to make smart those vintage bulbs that would have transformed his living room in a "Peakey Blinders" saloon-like-style. It probably did not occur to Torey that the socket design features were the result of **the choices and discussions** designers undertook while evaluating the best solution against at least two goals: to give customers a functional product and to hit the market.

Plugging objects in to the internet, ambiently sensing and enhancing people's capabilities with remote control are some of the possibilities afforded by the IoT. Yet its benefits are often marketed with little concern about the possible **threats** it may bring in terms of **unregulated collection and processing of data, lack of minimum security provisions**, etc. Understanding these aspects is pivotal as far as new technologies evolve faster than the legislator's ability to regulate them. For this reason, an approach that looks *ex-ante* at the consequences (by mapping the moral reasoning of those subjects who develop IoT solutions and services) might help to foresee the downsides that, if left to mere compliance with the law, will only be tackled *ex-post,* once they have already spread their effects on society.

In Virt-EU, we are deeply concerned with this issue as we aim to develop a framework that would foster **ethically-aligned design** among IoT developers. Thus, to exhaustively map the set of values that potentially drive designers' actions, our researchers form **London School of Economics** and **IT University of Copenhagen** used a set of participatory, ethnographic and interview methods across the events listed in the table below (figure 2). They further enlarged the mapping to IoT Manifestos

because "they represent a loud invitation to think in new ways[1]", delivering an oppositional perspective to the kind of moral reasoning expressed in larger IoT conferences.

Geographic location	Event	Date
Barcelona, Spain	Startupbootcamp	January 4, 2017
	4 Years From Now (4YFN)	February 27-March 2, 2017
	Mobile World Congress (MWC)	March 2, 2017
London, UK	Connected Seeds and Sensors	February 1, 2017
	Advisory Panel: The Impact on the Internet of Things on Managing Work (Loughborough University research project)	February 22, 2017
	Smart IoT London Conference	March 15-16, 2017
	MEETUP: IoT London	March 21, 2017 - ongoing
	ZAIZI 'Data Driven Government Roadshow'	March 23, 2017
	WiTT (Women in Telecoms and Tech) meeting: Insights from Mobile World Congress 2017	April 20, 2017
	Avren's World: IoT Networks Conference	May 23-24, 2017
	TechXLR8 Conference	June 12-15, 2017
	Expert workshop on citizen/consumer engagement with policy-making for Internet of Things	June 13, 2017
	IoT Trustmark: Day 1	June 16, 2017
	IoT Trustmark: Day 2	September 11, 2017
	Organised IoT Trustmark event	September 24, 2017
	FixFest	October 6, 2017
	The Ethics of Coding and the Human Algorithmic Condition: The Algorithmic Condition Workshop	October 10, 2017
	NetGain Partnership Event on Algorithmic Accountability	October 25, 2017

[1] Parent, M. (2001). The Poetics of the Manifesto. Newness and Nowness. In Manifesto. A Century of Isms. University of Nebraska Press, x–xxxi.

Berlin, Germany	Open IoT Studio, Mozilla	March 23-24, 2017
	ThingsCon Salon	March 23, 2017
Lyon, France	The IoT Showroom (SIDO)	April 5-6, 2017
Torino, Italy	Mini Maker Faire	May 27-28, 2017
Copenhagen, Denmark	The Internet of Green Things Festival	April 10, 2017
	The Things Network Hackathon	June 1, 2017
	ThingsCon Salon, TechFest	September 6, 2017
	Dowse Workshop	November 2, 2017
Geneva, Switzerland	IoT Week	June 4-7, 2017
Bled, Slovenia	Living Bits and Things	June 19-20, 2017
Malmö, Sweden	IoTConf.se	May 23-24, 2017

Figure 2 - Events

Same issue, different set of values

The Emberlight issue deals with the mismatch between the physical and the digital part of the product purchased by Torey. Mapping the **informal moral reasoning** regarding this problem revealed a different perspective that stems from the **double empirical exercise** we undertook.

In major IoT conferences, the issue of **"product durability"** was framed in terms of "**responsibility and design**". The subjects interviewed were mainly concerned about the recourse someone could take if an IoT company goes out of business and the end-user needs regular software updates to the IoT product they own. Furthermore, we observed that in the case of an IoT producer who outsources the software development to a third party, the latter may claim that if the product breaks, liability will rest on the manufacturer of the hardware and not on the developer of the software. According to this perspective, the problem concerning "product durability" is addressed in terms of who is responsible if Torey makes a legal recourse, once the smart socket stops to function. We have interpreted this widespread attitude as the consequence of pressure from market hegemonic forces. For instance, in the case of Emberlight, the choice to opt out for cloud routing seems to be grounded in the attempt to lower the cost of the smart socket, a necessary condition to survive the tough competition of other start-ups and bigger firms.

Instead, in the analysis of Manifestos, the durability of an IoT product was framed in terms of "**sustainability**". Because IoT products are made of hardware and software, usually the lifecycle of physical objects is longer than that of contemporary software, due to the fast rhythm of its continuous development. Several documents call for a re-alignment of digital and physical lifespans, addressing the problem of firmware and software updating that can often make perfectly functional hardware components unusable. The mismatch between lifespans of hardware and software often leads to intentional obsolescence of products. This kind of strategy looks to Manifestos writers as the **consequence** of the marketing decision to push an artificial demand for

new devices, something they strongly criticize. Instead, every IoT product should guarantee its functionality, as long as the hardware is still working. Following this standpoint, Emberlight designers shouldn't have relied on cloud routing as a central feature of their socket and Torey wouldn't have been left in the dark.

Studying the moral reasoning behind the design of IoT products shows that the same issues may be faced relying on very **different systems of values**, something that leads to extremely different solutions.

First insights and future research

The mapping of values in major IoT events returned a picture where **the application of law is viewed as a means to address ethical issues**. Law seems to work as a limit-case for ethics. As a matter of fact, for many IoT developers ethics is articulated with words also appearing in laws they must comply with. Only few of them point to ethical concerns beyond those commonly felt as relevant in the IoT domain, such as **privacy** and **security**. This is to say that very few developers are concerned with the problem of product durability faced by Torey, because for them a product like the Emberlight socket must be compliant with the laws, more than being designed to serve people needs.

In a few locales such as "ThingsCon Salons", "The IoT Trustmark" and "the Open IoT Studio Retreat", ethics is enacted as a matter beyond compliance with existing laws, something which is visible in discussions that also occasionally move on legal borders or point towards future laws. These events draw upon a wider perspective on ethics, similar to that coming out from the analysis of manifestos. Indeed, their authors wish to do their part in shaping technological development. In general terms, Manifesto writers seek to pursue a twofold intent: on one side, they suggest **future solutions to fix current problems related to technology**, while from another perspective **they pose deep concerns about technology itself**. The standpoint emerging from these documents seems to look at the design of IoT products and solutions with an ethical approach, suggesting that people who share that approach believe that decisions regarding the design of IoT products are the result of a process which must put moral reasoning at the center.

Our mapping has highlighted that a focus on these alternative positions, combined with data-driven identification of specific sites for sustained involvement within the most active IoT regional hubs across Europe, can help to provide a strategy for formulating our future engagement with this domain. As a matter of fact, we do intend to continue the engagement with these locales, because in the year to come we will enter the places where IoT products are imagined and designed.

In addition, to continue our fieldwork in the two sites we have selected as relevant *loci* for in-depth domain mapping, namely Amsterdam and London, we will also use network (link to articolo multilayer analysis) and legal research (see next blogpost) to support the identification of specific start-up partners who might become objects for a long-term ethnographic fieldwork.

BLOG POST 2

Data Ethics: Legal and Regulatory Aspects of Data Ethics

Miss Mabelle Bayard was Thomas F. Bayard's daughter, a United States Senator from Delaware and former candidate for President. When Mr. Samuel Warren married Miss Mabelle Bayard in Washington on January 25, 1883, he probably was not aware of the long-lasting attention the press would have bestowed upon him and his wife's family in the years to come. Indeed, such a noble family gave many stories to the scandalmonger newspapers that sought idle gossip. The constant attention of the press bothered Mr. Warren to the point of looking for a legal safeguard for his family private life.

Mr. Warren was a lawyer and therefore probably aware that the property right was considered as a safeguard against intrusion into a person's private life. This observation was borne by a well-established jurisprudence and articulated in the proverbial expression: "A man's house is his castle". It is attributed to Sir William Blackstone (1723-1780) (one of the fathers of the English Enlightenment) and for more than a century it stood as the bulwark for people's enjoyment of a private sphere. This private sphere was recognized as long as the threshold of physical private property was not overstepped.

While this legal principle lasted for a long time, it started to lose its ground when new technologies made the non-physical intrusion into a person's private sphere possible. The second half of the 19th century was a period of great technological advancement and innovation. Inventions such as the microphone and the camera made it possible to sneak into people's intimate life with no need to trespass someone's physical property. This led Mr. Warren and his colleague Louis Brandeis to look for new safeguards to protect people's "right to be let alone". In fact, in 1890 they published an article on the Harvard Law Review that laid the foundation for the long tradition of privacy and data protection doctrine. Their article, borne by an elegant rhetoric, tried **"to consider whether the existing law affords a principle which can properly be invoked to protect the privacy of the individual" [2][1].** They understood that when major technological advancements occur, well-established principles, even if affirmed by a robust set of judicial decisions, may soon become obsolete.

As the core objective of Virt-EU is to **identify ethical and social values and make them operational through the development of the Privacy, Ethical and Social Impact Assessment (PESIA) (link)** model, we have found it pivotal to run a legal analysis about the European regulatory framework on data protection. Such an effort has been spent in the willingness to develop the PESIA consistently with the existing regulation and principles. For this reason, **Polytechnic University of Turin** and **Open Rights Group** have been vested with a similar task to

---

[2][1] Warren, Samuel D., and Louis D. Brandeis. "The Right to Privacy." *Harvard Law Review* 4, no. 5 (1890): pp 197. doi:10.2307/1321160.

that accomplished by Warren and Brandeis in their seminal article: **i)** looking at the limits of the existing regulatory framework and **ii)** ascertaining which ethical and social issues in data processing are taken into account by DPA's, Article 29 Working Party, European Court of Human Rights, European Court of Justice and privacy practices.

The evolution of data protection regulation

Warren and Brandeis were fully aware that when major technological changes occur, the recognition of new rights is needed, since technological innovation runs faster than the legislators' capacity to hold technologies accountable for the effect they spread on society. Considering this, we performed the first task mentioned before exploring the evolution of the European regulatory framework in relation to the spread of new technologies.

Because of the challenges brought in the early 50's by governments and big corporations with the creation of large databases, where personal information could be aggregated, retrieved and connected to a wide extent, the legal dimension acted as a mere instrument to express and harmonise these issues, shaping data processing to tackle the risks of new forms of discrimination and societal control.

This cast the notion of data protection as the idea of **control over information**, thus the first data protection regulations gave individuals a sort of counter-control over collected data. As a matter of fact, legislators pursued this goal by increasing the level of transparency about data processing and safeguarding **the right to access to information**. Yet with the advent of personal computers in the mid-80s, new forms of marketing based on customer profiling and extensive data collection prompted legislators to focus on the economic value of personal information. As a result, citizens claimed more power in terms of negotiation against businesses exploiting their personal data. The **Directive 95/46/EC** represented the answer to this issue, introducing the "notice and consent" model.

Today we are experiencing an advancement in the analysis of large amounts of data collected from multiple sources, facilitated by the development of cloud computing and big data analytics. These technologies make it possible to **monitor social behaviours, infer patterns of behaviour and apply such patterns to individuals, to predict their actions and take decisions that affect them, which might lead to discriminatory practices**.

In this context, it is furthermore essential to recognize the nature of the fundamental right of data protection (Art. 8, EU Charter of Fundamental Rights), to create a barrier against the commodification and reification of personal information. This is probably the main contribution, in general terms, of the legal framework: to harmonize societal values and the use of data to develop new devices.

The challenges brought by the IoT and why the law alone does not suffice to address them

Returning to the story of Mr. Warren, what particularly bothered him was the meticulous description newspapers gave about the parties his wife used to host, the colour of her dress, the people who participated: this was the kind of information he would have kept private. As Mr. Warren was aware of the purpose behind the exploitation of this information, he had no reason to fear an intrusion beyond those situated events.

In this regard, the IoT is a game changer. Since individuals voluntarily choose to introduce smart objects into their houses, someone might reasonably argue that they are acquainted with the hidden practices performed by the objects themselves: how the data gathered are analysed, the purposes of their processing, the kind of technology involved, the categories of data processed and the eventual imbalance of power between the data subject and the data controller. These are the issues that should matter nowadays to citizens.

Yet the ubiquitous and invisible nature of IoT devices makes it difficult to understand to what extent the hidden practices they perform affect people's daily life. In her book, Virginia Eubanks (Automating Inequalities) observes that *"many of the devices that collect our information and monitor our actions are inscrutable, invisible pieces of code. They are embedded in social media interactions, flow through applications for government services, envelop every product we try or buy. They are so deeply woven into the fabric of social life that, most of the time, we don't even notice we are being watched and analysed."* [3][2] This is due to the fact that the IoT is everywhere: we dress it (**Body Area Network,** e.g. wearable devices) we welcome it in our homes (**Local Area Network,** e.g. smart home appliances), we inhabit the cities where every step we take can be potentially registered and analysed at fast and growing pace (**Very Wide Area Network,** e.g. smart city). So, what if the seamless flow of information across these different dimensions were merged and analysed without restrictions? Such a possibility is rendered even more troublesome by **machine learning algorithms**, as they pose new problems by reducing human intervention in the processing of personal data, increasing societal issues regarding decision-making processes. For example, an insurance company may deny the insurance coverage due to an error in the processing of clients' data. Or even worse, it may set the price based on the social group we belong to.

In this regard, law does not suffice to grant citizens adequate safeguards, because while protecting common values such as privacy, personal identity and dignity, it falls short to address those ethical and societal issues ignited by the advent of the so called "algorithmic society". It works well in protecting the individual, but it lacks the capacity to tackle those decisions based on obscure algorithms that end up being more biased towards certain social groups.

---

[3][2] Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, St. Martin's Press, 2018.

Thus, how can we go beyond these limits? Should we look, as Mr. Warren did, at a court's ruling, waiting for the establishment of a solid jurisprudence from which to deduce new rights? Or should we rely on a strong regulation that prevents the aforementioned issues to occur?

European Union's approach towards data protection

In this part of the project, **Politecnico di Torino** and **Open Rights Group** analysed the regulatory mix present in Europe (i.e. data protection legislation, judicial decisions, guidelines, charters of values, best practices and standards), to understand if it takes into account the social and moral values threatened by the rapid technological development. The picture that emerged shows that while the cornerstone balancing of interests achieved in the **Directive 95/46/EC** tried to satisfy the demand of data subjects to be in control of their data (recognising a prominent role to individual consent and imposing information duties on data controllers), over the years the **"notice and consent"** mechanism has showed its limits in providing an effective safeguard to moral and social values.

Such an insight has been further confirmed by the empirical analysis based on the **2015 Eurobarometer on Data protection** and on the **2016 Flash Eurobarometer on e-Privacy.** According to these surveys, the notice and consent mechanism has fallen short to bestow people with sufficient control over their data: indeed, only a minority of data subjects fully read privacy policies. Furthermore, it is inadequate, given that people are largely unaware of the kind of their data collected and analysed.

This is an important element that should be considered in the following stages of our investigation (Deliverable 4.1). As a matter of fact, we will conduct an in-depth analysis to assess whether the GDPR offers new solutions to bolster data subject's interests. Indeed, our aim is to ascertain if the upcoming General Regulation affords protection to people for what concerns the collective dimension of data processing and the consequent risks it brings. In this vein, we will see how the data protection framework will be forged by the complex interplay between the requirements present in the GDPR and the **future judicial decisions** that will be issued by EU Courts (European Court of Human Rights and European Court of Justice), national data protection authorities and Art. 29 Working Party. Thus, a mixed approach that foresees the integration of the GDPR with a set of judicial decisions is the solution we follow as it is more apt to grant citizens a legal safeguard against the unrestricted use of data.

First conclusions and further investigation

Based on these findings, it seems that the regulatory mix developed in Europe since 1995 on data protection is only theoretically able to take into account the social and legal implications of data uses. Values are often implicitly considered by the different components of the regulatory mix, but there is a lack of tools which can make values explicit and operationalise them.

This does not mean that the regulatory mix ignores the importance of ethical and social values, but it implies that it has difficulties in putting them into practice in a clear and direct manner. In this regard, the **PESIA** will represent a concrete tool in the hands of developers and designers to assess the risks at stake with data intensive practices.

BLOG POST 3

Considering ethical and social values for a better accountability: the PESIA Model.

When Alex bought the XFace video cam, he was probably unaware that what unlocks his house door with no need to insert the key is a face recognition algorithm powered by artificial intelligence.

The main reason that led Alex to opt for a keyless system was to overcome the countless occasions he had been left out of his house because he had forgotten the key. Moreover, XFace was relatively cheap and came with many features that made it a best-buy product.

By purchasing it, Alex agreed to the **terms and conditions** the producer issued as a requirement to process the data collected by the camera. But because Alex was not a computer scientist or an expert in artificial intelligence, he was probably ignorant about how the software that powers the camera works. Code libraries and neural networks trained to process million pictures are complex topics that are neither at hand of a layperson nor accessible by reading the full terms. Alex ignored, for example, that some of the most common libraries to power face recognition algorithms are exposed to **gender** and **racial bias,** as proven by MIT's researcher Joy Buolamwini. Indeed, some of these libraries outperform when it comes to "lighter-skinned women, but err at least 10 times more frequently when examining photos of dark-skinned women".

This represents a risk whose consequences had not fully been taken into consideration by Alex when he left the "decision" on how to let him enter his house to IoT designers .

An explanation to figure out Alex's decision is that many human activities rely on a different perception of risk. Indeed, the risk of an undesirable event can be measured in quantitative terms by its **probability** and its **severity**. The first concerns the chance that the undesirable event is to happen - in this case, being left out because of an algorithm bias - while the second concerns the size and seriousness of the consequences - for example, Alex's young daughter is in danger and he cannot enter. The combination of these two dimensions leads individuals to evaluate the acceptability of risk in relation to the expected consequences that may follow a decision.

However, there are some activities whose risk is difficult to be assessed. Among them, those regarding new technologies that collect and process a huge amount of data are particularly relevant. It is the case, for example, of the IoT whose ubiquitous nature raises major concerns for its massive collection of data.

The Eclipse 2018 IoT Dev survey revealed that "developers are starting to realize that beyond the "cool" factor of building connected devices, the real motivation and business opportunity for IoT is in collecting data and making sense out of it". Seamlessly gathered IoT data fuel machine learning-backed-decision-processes that are gaining momentum in people's daily lives. This is true inasmuch as these techniques assign them to certain clusters that determine, for example, how goods and services are supplied to their owners and customers.

Far from refraining developers to engage in new business opportunities, the General Data Protection Regulation (GDPR) requires those who are responsible for the collection and processing of data to address the risk bound to these activities. Starting on May 25th, the GDPR (**which has**

unified the EU's 28 members regime of data protection) has introduced under article 35 a procedure to assess this risk.

Our project is deeply concerned with the development of a risk assessment tool (**PESIA**) that would help developers and designers to assess the societal and ethical risks bound to certain design choices. For this reason, Virt-EU researchers from **Politecnico di Torino** have gone in **deliverable 4.1** through an in-depth analysis of the risk assessment model foreseen by the GDPR, highlighting its strengths and its weaknesses. This analysis has been done in the willingness to develop a tool that is consistent with the GDPR's Data Protection Impact Assessment, but at the same time strengthens the capacity to assess the ethical and societal risks bound to the processing of non-personal data.

The risk assessment rationale in EU before the GDPR.

Touted as a revolution of the data protection discipline, the GDPR has been welcomed by worrisome headlines all around the web (here, here). Yet, beyond the noisy claims, what the GDPR basically does is to rethink the risk assessment rationale.

In doing so, the legislators revived the prominence of the **accountability principle** set forth by the Council of Europe in **Convention 108**. The rationale behind this principle was to hold the data controllers responsible by asking to put in place all the adequate measures to guarantee a lawful processing of data. In other words, the **data controllers must in the first place be responsible for the assessment of risk** coming with the processing.

On the opposite, the diffusion of computers registered in the eighties led legislators to pursue another orientation. The idea was to put on data subject's shoulders the burden to self-assess the consequences of data processing. The diffusion of computer among the masses was thought to enhance greater individual awareness regarding the electronic processing of information. For this reason, through the **Directive 95/46/EC**, legislators sought to put the emphasis on individual decisions, transforming accountability in "**terms and conditions**". This rationale was prompted by the consideration that to the extent data are important to shape personality and individual life: the best judge to run the process of informational self-determination was the individual herself.
Thus the "notice and consent" mechanism as featured in the **Directive 95/46/EC** was inclined to an individual assessment of risk.

Today's technological landscape may lead someone to label legislators' choice as naïve. Yet they could not have foreseen that individuals' informational self-determination would have been swallowed up into (or by) a **data maelstrom**. As a matter of fact, the advent of **machine learning techniques** and the **proliferation of data sources** have made it possible to unitize, swirl and cross-check data that are increasingly used to fuel **automated decision-making processes**. In such a realm, data subject's right to informational self-determination is partially stricken off.

For this very reason, while preserving the individual consent mechanism, the GDPR has gone into the direction of a tougher accountability. As it is featured in the GDPR, the model of risk assessment draws upon an array of procedures and principles that go into the direction of a stricter assessment to be performed by those subjects concerned with the processing of data.

How should a data controller assess the risk of data processing within the GDPR framework?

General Principle

What does it mean to be responsible and to thoroughly assess the risky activities bound to the collection and processing of data subject's data in the aftermath of GDPR's entry into force?

The example of XFace video cam might help answer this question.

Along with the face recognition algorithm, XFace video cam comes with other functionalities. It allows customers to get a message every time a car is parked in their property and at the same time a video is live streamed to their phone. The camera installed inside the house is further featured with a built-in control system that checks if light bulbs are switched on, thus allowing customers to remotely switch them off.

Now XFace video cam has to comply with the GDPR, which imposes a **right-based approach** to risk assessment. It doesn't consider a risk in terms of a **tradeoff between risks and benefits**, but it preserves some fundamental rights: as a matter of fact, when it comes to data protection, every risk that could damage a data subject's right should necessarily be avoided, no matter what is lost in term of benefits.

For this very reason, the first consideration is that data controllers should be mindful about the processing of personal data. Art. 4 recalls the definition of personal data as "**any information relating to an identified or identifiable natural person**…".
Thus, the controllers should primarily focus their attention towards those data that, if processed, might cause "**material and non-material damages**" that prejudice the "**rights and freedom of natural persons**" (Recital no. 75, GDPR).

Data collected through the XFace recognition algorithm surely belong to this category, as well as location data, online identifiers, identification numbers, that can all be used to indirectly reveal someone's identity. Once the type of data at the heart of the processing operations is ascertained, the data controllers have to deal with the "**purpose limitation**" and "**data minimization**" principles (art. 5). Before starting any assessment of the risk, the controller should limit the processing to those data that are necessary to deliver the service for which the data have been collected. In the case of XFace, the data collection regarding light bulbs has to be limited to the feature of remotely switching the light off. Extending the collection of data to infer energy consumption patterns may go beyond the consent for processing to which data subjects have agreed. Meeting these principles constitutes a precondition to start the risk assessment.

The risk assessment model in the GDPR

The risk assessment model is mainly enshrined in art 24, 32, 35 and 36 of the GDPR.

The general requirements expressed through the purpose limitation and the data minimization principles are devised to help XFace ascertain whether the processing of personal data comes with some risk. It is then complemented by a set of measures listed in art. 32, that aim to implement by

default all the **technical and organizational solutions to minimize the impact of data use on individual rights and freedom** (e.g. pseudonymization, anonymization, limits to data retention).

The assessment procedure is called for in the wording of art 35 "**Data protection impact assessment**". The procedure entails for data controllers such as XFace to perform an impact assessment based on different modules. Its modularity is outlined in art. 35.7. and ideally foresees the following steps:
i) *to make a recognition of the processing operations and of the purposes for processing*;
ii) *to assess the necessity and proportionality of the processing operations in relation to the purposes*;
iii) *to assess how the risks might harm the right of the data subjects*;
iv) *to select and implement those measures to prevent or mitigate the risks.*

This scalable model sets a threshold in the notion of "**high risk**" to the rights and freedom of natural persons, yet not providing a clear definition of high risk. Instead, in art. 35.3 three cases are specified in which the DPIA is required. What emerges from these observations is the non-mandatory nature of the DPIA, though the national data protection authorities can adopt either a list of cases in which DPIA should be performed (art. 35.4) or a list of cases in which a DPIA is not required (art. 35.5).

Even though all the steps foreseen in the assessment procedure have been carried out, if a high risk still persists, data controllers can ask for help from the national data protection authority for a prior consultation (art. 36).

The limits of the assessment model

We have seen how the model of risk assessment as featured in the GDPR has been thought to evaluate the risk linked to the processing of personal data. But how does it perform with those data that are not personal, yet can be used to take decisions that discriminate against the social group Alex belongs to, thus indirectly affecting Alex as an individual?

In the age of big data, everyone should be mindful that the more services someone gets from a technology, the more data are collected for processing. Such an observation is heuristically useful to observe the kind of data XFace video cam may collect beyond those considered personal in nature.

For instance, every time a member of Alex's family enters the house, the face recognition software registers a timestamp value which can be used to reconstruct the habits of the family: who works, who stays at home most of the time, and based on the number of family members it can infer how many children Alex has.

These data, enriched and cross-checked with other data sources (e.g. **census data**), can be used for **predictive policing** that may lead to **discriminatory practices**. For example, let's imagine that Alex is living in the outskirt of a big city. Maybe he moved with his family because, for the same rent he used to pay in the center, now he can afford a bigger house. His wife works from home as a freelance journalist. The family is complemented by his young son and his older daughter frequenting the high school. XFace recognizes that the family is composed of four members and that while three of them every morning get out from the house, a member spends most of her time

inside as commonly does a housewife. If these data were cross checked with those that show that the zone where Alex's family lives is a low-income area and that his wife is probably a housewife, a credit scoring system might infer that the family is not eligible for a loan because it relies on a single salary.

It should be noted that these forms of discrimination are not necessarily against the law, especially when they are not based on individual profiles and only indirectly affect individuals as part of a category, without their direct identification. Moreover, within the EU, such data analysis focusing on clustered individuals may not represent a form of personal data processing, since the categorical analytics methodology does not necessarily make it possible to identify a person.

The PESIA model

The aim of the PESIA model is to develop an assessment tool that will help to pay greater attention to ethical and social implications of data use, as we have seen in the example of Alex. For this reason, we are working to develop an agile tool to be used on a voluntary basis. Furthermore, we wish to promote an open and participatory approach to risk assessment (**DPIA is internal and not public**).

The ambitious objective we defined by devising the development of such an assessment tool is therefore the consideration of those values that go beyond those protected under the GDPR (right to data protection, security, integrity of data, etc.).

So, the model we have foreseen to overcome the lack of accountability to social and ethical values is featured with three different layers:

1) the common ethical values recognized by international charters of human rights and fundamental freedoms;
2) the context dependent nature of the values and social interests of given communities;
3) a more specific set of values figured out by IoT developers, concerning the specific data processing
application..

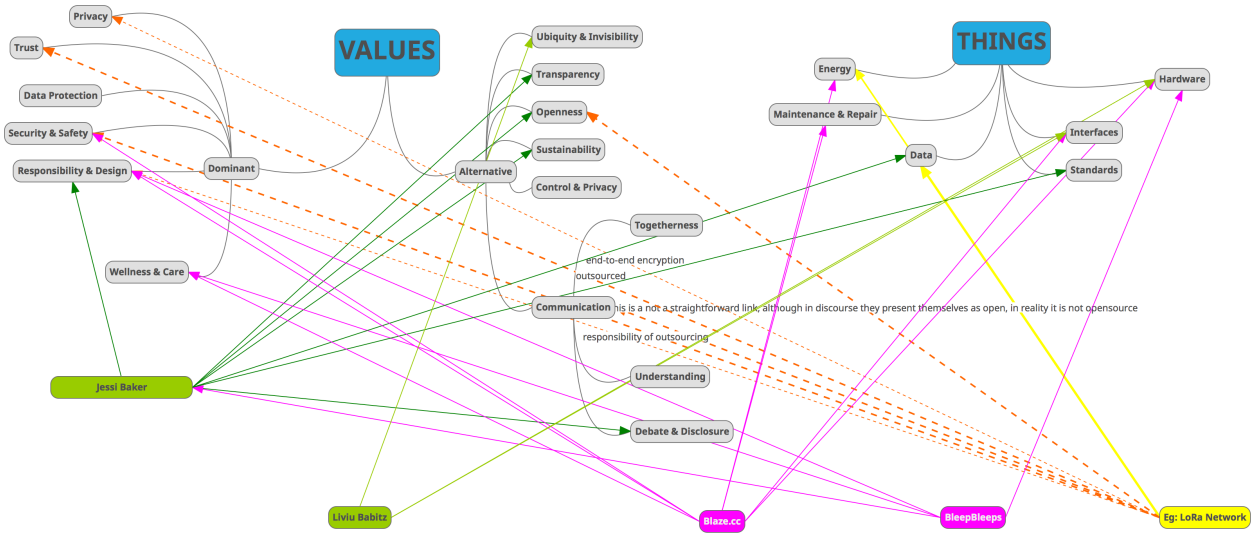
The main aspects outlined in this article suggest that the existing Data Protection Impact Assessment should evolve into a broader and more complex Privacy, Ethical and Social Impact Assessment (PESIA). We are strongly committed to develop this tool in the willingness to foster an **ethical attitude** towards Europe's policymakers, industry and developers' communities.

SECTION 2: Mappings

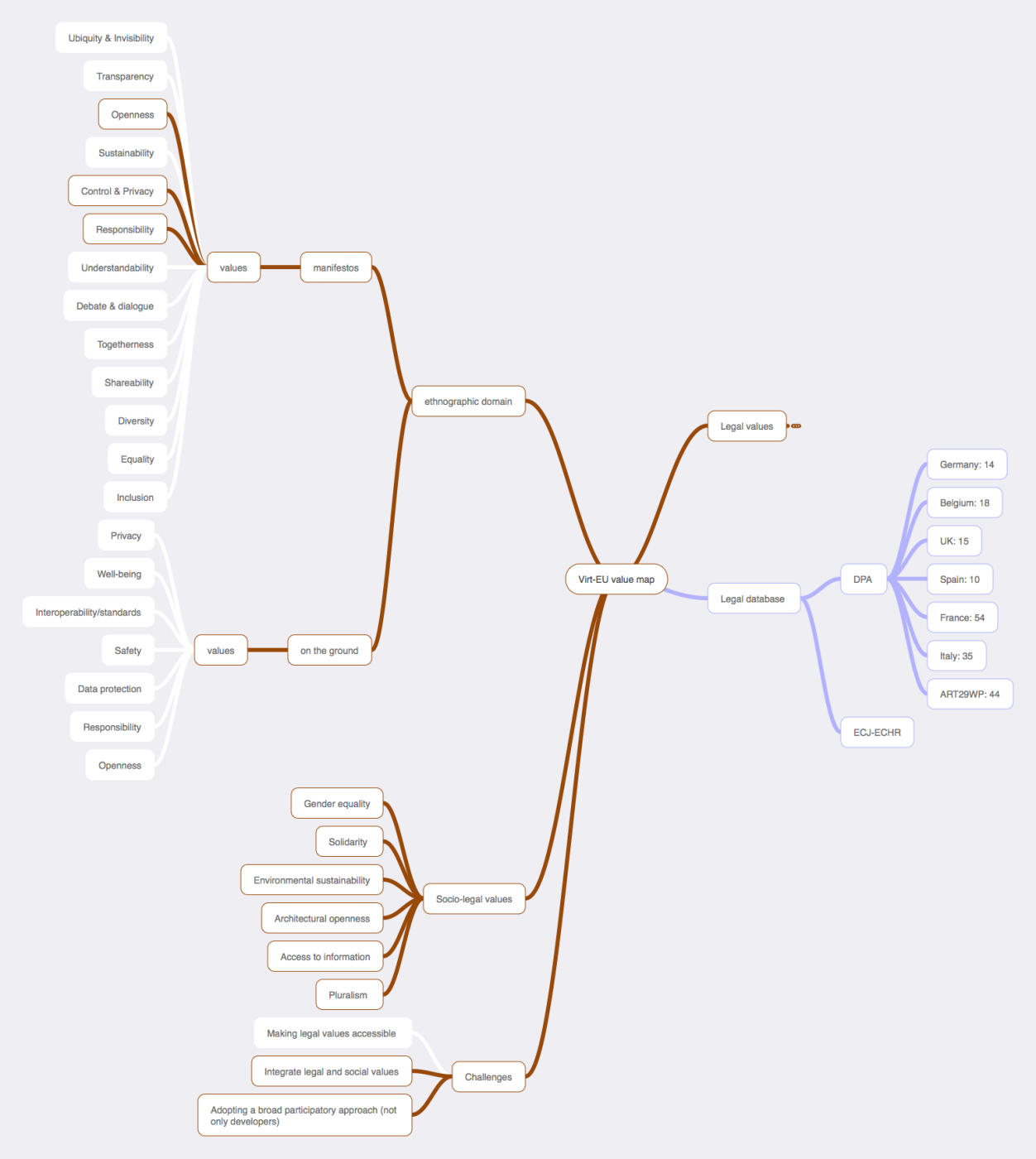
In order to communicate more meaningfully and effectively both within our project consortium, with our community of co-designing developers, as well as with the general public, we have begun to compose a series of mappings and diagrams to demonstrate otherwise difficult to describe connections and constructions. These mappings and diagrams will form the basis for an interactive media visualization of values coming from different sources that will be made available as an interactive artifact towards the end of the project. The reason for not making this interactive visualization sooner is in order to avoid formation of presumptions and expectations among the developers that we continue to engage.

Section 2.1: Legal + ethnographic value-mapping

The first two maps are focused specifically on value mapping: from the ethnographic fieldwork to the legal analysis work. As is evident, certain values uncovered through ethnographic fieldwork connect to certain values identified by the legal team. However, some are not connected. The gaps we can identify as well as overlaps both provide foundations for the content of the tools that CIID will prototype through co-design sessions.



Map 1: Values identified during ethnographic fieldwork
Source: LSE + ITU Presentation, May, 2018



Map 2: Values identified through legal analysis combined with ethnographic fieldwork discoveries
Source: Politecnico di Torino, PESIA Presentation, May 2018

This series of maps from the partnerships collaborations - specifically between LSE, ITU, and Politecnico di Torino - demonstrates the evolution and sharing of values identified during ethnographic fieldwork and research as well as socio-legal analysis. As is evident, certain vales, whether they be dominant or alternative, occur in both the ethnographic mapping of findings as well as the socio-legal mapping of findings. If we follow the value map (Map 2) from top to bottom, we see that certain values emerge in the manifestos of the IOT companies in study (see blog post 1 for more details about the analysis and collection of the manifestos). While some values from the manifestos are the same as those that can be found in the "on the ground" study, such as "Responsibility", this exact word is not immediately, obviously present in the socio-legal values

map. However, as the ethnographic research team is continuing to further the way that they define each of these values, it is likely that we will start to see how the ethnographic domain values do in fact overlap with many of the socio-legal values. Already, we see at least one direct overlap: sustainability and environmental sustainability.

Once we have laid out and digested the various similarities and differences between the values surfaced by two branches of the project (ethnographic and socio-legal), we can build tools for ethical reflection that would take into account both branches. Why are values important to identify and map out at all? As described in the manifestos, "The mapping of values in major IoT events returned a picture where the application of law is viewed as a means to address ethical issues." Thus why would it be important to search more deeply for other articulations of values? As discussed in "Searching for moral reasoning in the IOT" (Blog Post 2), there are smaller decisions made throughout the design process (such as choosing to use the bluetooth protocol that would store data on a third-party server instead of in a local server). These small decisions have trade-offs - good and bad consequences - and the trade-offs may be acknowledged in the moment but are not necessarily reflected in a broader way unless they add up to a more blatant disregard for the law. However, if the small decisions and more informal moral reasoning that occurs during the process of designing IOT were to be considered as just as relevant for clear articulation, and merit critical questioning - because these are moments where a start-up could shift its path just slightly and avoid a bigger problem in the future. The tools that we co-design with developers will work towards making values explicit and operationalising them. The map developed by LSE (Map 1) already gestures at the potential of literally tying values to “Things”, according to our community of developers and designers, and this is an insight that we are pushing farther into the prototyping process of tools for ethical reflection. That is - if an IOT developer identifies a given value as one of their core ethical values (for example, Responsibility & Design), and considers how that value is often tagged to be relevant to the treatment of Data, this sort of articulation (in the previous two steps) could lead to a thorough questioning of how the IOT developer is working with data in the product at that moment.

As the following diagrams on ethical approaches show, our project uses several ethical frameworks and approaches to help developers and designers conceptualise their points of view and understanding of their work in relation to their ethical reasoning. This means that we will consistently probe at what IOT developers and designers think of as "Good". As the diagrams show, this notion of "Good" is filled in with more descriptive values based on the community around the IOT developer - a community that is also influenced by the overall socio-legal values as presented in the map above. Thus through these mappings and diagrams, we should start to see how a sense of ethics can be formed with the foundation of both ethnographic and socio-legal values, and then questioned through the application of a series of different lenses through which we might look at a given problem.

Each of these diagrams represents the outcomes of a series of conversations and interviews between CIID and ITU. We created the diagrams not only to attempt to visualise a system that can sometimes go beyond words, but also to be able to more succinctly communicate the ethical approaches our project considers to the co-design partners we have been working with for the co-design workshops where we come up with tools for ethical reflection during the IOT design process. Given that our co-design partners do not have time to read the many books and papers that support and explain each approach, we needed to create diagrams that would provide the basis for

our conversation and brainstorming around what we mean by ethics when we consider ethics and IOT.

Section 2.2: Diagrams of ethical approaches

The following diagrams present different mappings of the ethical approaches that the ethnographic research from LSE and ITU has identified as relevant in terms of the overall project's awareness and presentation of what we mean when we talk about ethics.

The first diagram presents a possible interactive concept in terms of being able to mouseover and learn more about different areas of the otherwise complex diagram. As the project continues, we will continue to redesign and evaluate how we might present such complicating ethical concepts through visualisation and interactive multimedia techniques.

The following diagrams are a result of interviewing and sketching between Irina Shklovski of ITU and CIID's research and design teams. While blog posts and articles will be relevant for many of our audience, the series of interactive diagrams will serve to help our co-design community, stakeholders and broader public to slowly immerse themselves in the complex theories of ethics that they genuinely want to understand. As they immerse themselves in the different aspects of these theories, they will also find how our tools for ethical reflection might augment one element or another. Therefore, they (our co-designers and stakeholders) will also have the core content of the project as material for re-imagining the prototypes that we will be sharing with them. One possible implementation of the diagrams as interactive experiences would be a simple mouseover experience. Another could be uncovering different snippets of a podcast-type audio story as we move across the diagram (in a cinematography style). These different ideas for presentation and experience will be refined and implemented over the course of the next several months as we continue to push our project to distil its findings into tangible forms that more of the general public could experience and understand.

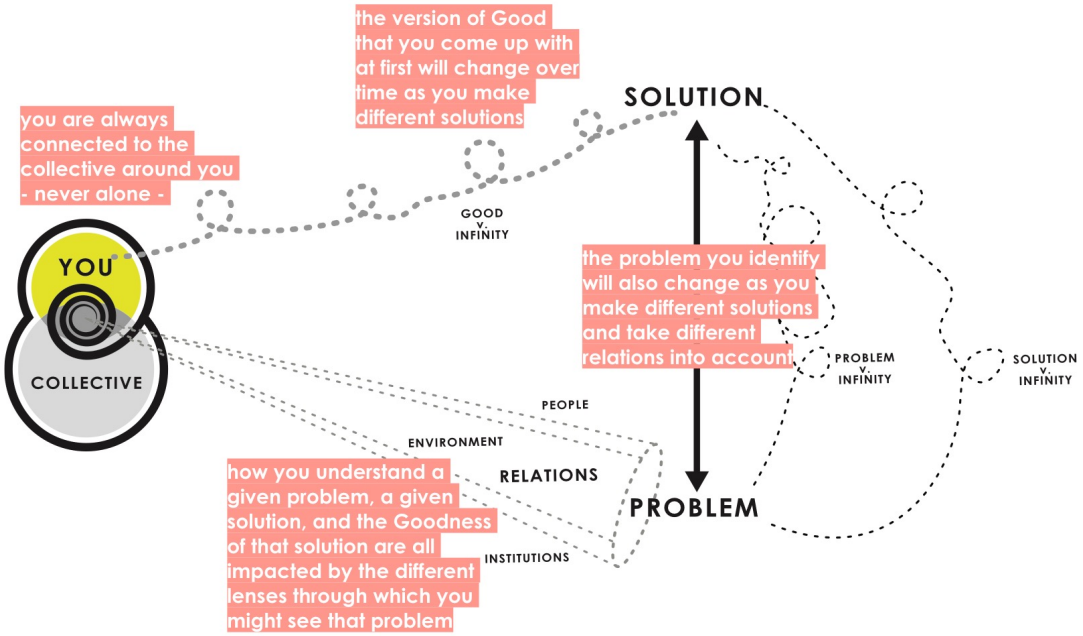


Diagram 1: Care Ethics, annotated

In the Care Ethics diagram, as shown through the key aspects are pulled out through annotations - this gestures at the next iteration of these diagrams into an interactive multimedia experience. Here we demonstrate that each piece of information that is related to the different parts of the diagram could be revealed through a mouseover or click, for example. We will work through the presentation and experience of these approaches by the end of the year as we continue to display them in our co-design workshops and improve the experience of understanding how ethics relates to the design process. Thus in this diagram we see that first of all, "you" are always part of the collective - you are never alone in your decision-making. Secondly, how you see a given problem depends on the environment around you, the people who influence your view of the world, your relations, in short. Next, the solution you will come up with for that problem depends on your notion of what is "Good." Your notion of what is "Good" will also consistently iterate as you come across different relations, different problems, different solutions. Therefore we have three "v. infinity" loops - for the notion of "Good", the problem you will see, and the solution you might envision. The care ethics is an infinitely looping approach of trying to understand the invisible relations that are connected to a problem, of trying to come up with a solution for that problem, and more than anything, trying again if your solution did not work.

As our partners at LSE and ITU write, in using the care ethics approach, we must "not only examine responsibility and care but take into account the shifting obligations and responsibilities of individuals as they are positioned in a web of relations. In our work, we are interested in the tensions between how individuals must negotiate their, at times conflicting obligations and responsibilities to others, and how they are expected to behave virtuously or 'well' in relation to a

ideal set of future potential states of being. How then must we consider what constitutes “doing good” given the conflicting relational demands from team members, management, other institutional arrangements, personal relationships, diverse community memberships as well as from the moral objects of hardware, data and code? But the logic of care has no real use for guilt, because it merely calls for acknowledging problems and trying again. In this way, the logic of care offers a way around the paralyzing realizations of downright apocalyptic possibilities of IoT. Where might we seek solutions to these problems? Julie E. Cohen proposes the idea of "semantic discontinuity” as the opposite of seamlessness - a call for strategically under-designing technologies in order to allow spaces for experimentation and play. Such intentional building in of flexibility may be one way to offer possibilities for alternatives.”

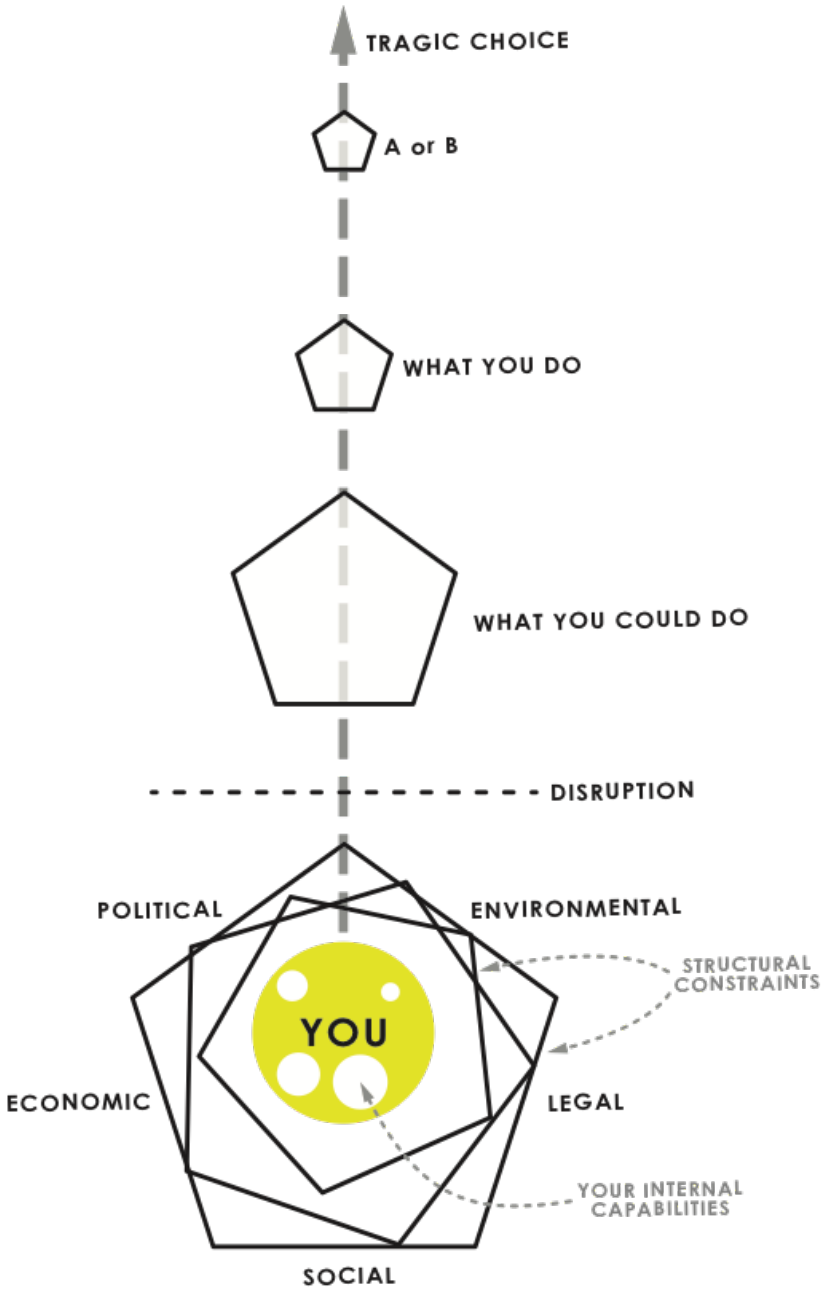


Diagram 2: The Capabilities Approach

The Capabilities Approach diagram presents the ever-narrowing constraints imposed upon individuals as they try to make ethical choices. So, in fact, when beginning to work on an IOT product, an individual might think they can align perfectly with their sense of ethics - with the values that they hold dear - the possibilities they practically face become more and more narrowly defined because of the combination of internal capabilities and the structural conditions defined by the particular social, economic and political environment within which the individual attempts to act. This recognition that personal principles may need to be compromised to cope with structural constraints point to the importance of understanding what these constraints are and what influence they might exert. Furthermore, technology developers are in a curious position of both having to make decisions within the structural constraints of their context and having to acknowledge that the design decisions they make will result in producing structural constraints and possibilities for their users. Thus for developers to "do good" it is important to not only evaluate how existing constraints affect design but also to consider how these constraints are translated into the design and how these might be mitigated to offer more or different possibilities to the users.

Here we also share the result of interviewing and compiling writing from LSE and ITU in defining this ethics approach in more accessible language:

"Choosing an alternative given the existing structural constraints and opportunities. 'Capability is thus a kind of freedom to achieve alternative functioning combinations.' This means that paying to attention to individual's internal capabilities is insufficient and we must also consider the possibilities created by a combination of internal capabilities and the structural conditions defined by the particular social, economic and political environment within which the individual attempts to act. This recognition that personal principles may need to be compromised to cope with structural constraints point to the importance of understanding what these constraints are and what influence they might exert. Furthermore, technology developers are in a curious position of both having to make decisions within the structural constraints of their context and having to acknowledge that the design decisions they make will result in producing structural constraints and possibilities for their users. Thus for developers to "do good" it is important to not only evaluate how existing constraints affect design but also to consider how these constraints are translated into the design and how these might be mitigated to offer more or different possibilities to the users."

Diagram 3: Virtue Ethics

In the virtue diagram, we see how you are constantly striving towards a notion of Goodness, and in this striving you are becoming virtuous. A good or bad outcome, a better or worse life, hinges on individuals actively cultivating particular virtues in themselves resulting in the kind of moral character that would lead to decisions with good outcomes. It is only through the experiences you will have as you strive towards the notion of Good that you gain practical wisdom. The last crucial aspect of this theory and slice of the theory into the diagrams the idea that your community and your relations will form your understanding of What Is Good. Thus despite this focus on the internal worlds of individuals, virtue ethics also emphasizes the importance of community. Virtue ethics gives most importance to the individual as an ethical agent in their decisions and practices and as a part of a community.

The basic definition of our VIRTEU project has moved from simply considering virtue ethics to also considering the Capabilities Approach and the Care Ethics approach because through the observations on the ground and workshops we have conducted with IOT developers, it is clear that the relatively direct Virtue Ethics approach is not sufficient to cover the myriad of moral reasoning and questioning that the IOT developers engage in on a daily basis.

Again, we share a short description based on our interviews with our partners at LSE and ITU to go with the virtue ethics diagram:

“An individual’s process of attempting to live a good life. Virtue ethics offers an individualist approach that sits well with the ethos of technological development, focused as it is on augmenting and improving the self. The familiar rhetorical devices such as “technologies for good” or “don’t be evil” speak to the idea that the virtuous moral choices of technology developers and designers can lead to bringing about a better life for all. From a virtue ethics point of view, such an outcome hinges on individuals actively cultivating particular virtues in themselves resulting in the kind of

moral character that would lead to decisions with good outcomes. Despite this focus on the internal worlds of individuals, virtue ethics also emphasizes the importance of community. Virtue ethics gives most importance to the individual as an ethical agent in their decisions and practices and as a part of a community."

Conclusion

The blog posts will be shared and validated by the partnership, used on the web-site, blog and social media. The preliminary value mappings as well as diagrams of ethical approaches will be considered during the next phase of production of interactive multimedia materials and will form the basis of artefacts for presenting the findings for the overall project.

In terms of the content presented and digested in the above mappings and diagrams, with the mapping of values and continued work towards defining the varying meanings of each value, we are making progress towards a clearer understanding of a basis of the notion of "What is Good" at least from the community standpoint that is surrounding an IOT developer in any given one of our field sites. As the blog post on "Searching for moral reasoning in the IoT" shows, each word has different associations with different overarching values - such as the issue of "product durability" being framed as relevant to "responsibility and design" and "sustainability" depending on the context. This means that we need to take some steps forward in the next two months in terms of either a) being transparent about the multiple definitions and leanings of these soft values or b) providing various definitions so that we can track how an individual IOT developer or designer is aligning themselves with different overarching values depending on which definition they choose. Furthermore, we will continue to push and develop our observations and understanding of how different values are linked to different "Things" (per Map 2). An early concept for prototyping has to do with raising certain warning signs when a given value is stated along with a given material choice - these may push the product towards a place of conflict where a value (for example, openness) is in contradiction with a material choice (for example, semi-open API), and therefore again, throws another value into consideration: "responsibility". It is clear that the web of values, choices and materials are in a constantly shifting pattern of definition. It will be up to our teams of ethnographic researchers and co-designers to ensure that our tools for ethical reflection incorporate the notion of values and "Good" as a moving target - that constantly evolves and changes. Each start-up or small company might bring their own slightly different interpretation of what a given value means in practice, for them, and our tools need to not only allow room for that evolution but also in fact encourage that evolution as something that can be driven by the community who will use our tools themselves.

With the presentation of ethical approaches as different diagrams and flows, we also open up the possibility of questioning how the design process itself could or should shift to allow for the integration of various moments of ethical reflection and self-assessment. If, for example, the notion of becoming good is an evolution that should happen over the course of an individual's entire lifetime, how might we capture the moments of evolution that happen over the course of an individual's work on a given IOT product? If through the lens of the capabilities approach, we notice how a team might struggle to shift a product's direction towards their notion of a Good direction (more responsible, more sustainable, for example), what kind of tools could we design that would allow for this acknowledgment to move from simple acknowledgment to possible action or

re-design or consideration of material changes within the product itself. In both of these approaches, let alone the care ethics approach, we have a consistent question which is when might or should a given type of ethical questioning be best fitted to a given design phase or moment in the design process? We will address this question through our upcoming co-design workshops as well as through key research on different design processes that our community of IOT developers and designers follow.