INCEPTION REPORT

RETHINKING INNOVATION. TOGETHER.

Andrea Bandelli - Brandon O'Dell - Catherine Hankins
Amsterdam Health and Technololgy Institute - November 2015

TABLE OF CONTENTS

RRI AND TECHNOLOGY SHIFTS IN HEALTH AND MEDICINE

PURPOSE OF THE REPORT

INTRODUCTION

''Sparks - Rethinking innovation. Together'' is a project to familiarize and engage European citizens with the concept and practice of Responsible Research and Innovation (RRI). Sparks brings together organizations such as science centres and museums, universities, research centres, science shops, and local administrative entities to produce travelling exhibitions, public events, and policy recommendations on how to best communicate about RRI with citizens. Sparks received funding from the European Commission under the framework program Horizon 2020 and runs from 2015 until 2018.

The project focuses on a topic of inherent interest for everyone: ''technology shifts in health and medicine''. Using examples and case studies from health and medicine, Sparks engages the public in conversations about the mutual and shared responsibilities of all parties that are active in scientific research and innovation. Sparks aims to facilitate discussions of the complex issues surrounding contemporary science in health and medicine, topics that are of immediate particular relevance for the public. It also aims to illustrate RRI in action.

The goal of this report is to offer to all the partners of Sparks a common understanding of the technological shifts that are influencing the fields of health and medicine and of how RRI can be identified, discussed, and modelled in the activities that will be developed for Sparks. The report provides an overview of current and future trends at the intersection of technology, health, and medicine. These trends are contextualized in the practice of Responsible Research and Innovation (RRI). Thus the report provides descriptions of the content that local organizers of Sparks will come across during the project, along with inspiration and examples on how to identify additional content items, develop storylines, and engage stakeholders for the development of local exhibition components and programs. The whole domain of technology shifts in health and medicine is so vast that it is impossible to create a complete catalogue of what is at stake. Moreover, developments in technology are occurring so quickly that by the time this report is published, and certainly during the lifetime of Sparks, new findings and innovations will disrupt the current status quo and open new horizons for scientific research and healthcare implementation. Therefore, we have taken a broad approach to define areas where technology shifts are more likely to influence current practices in health, medicine, and healthcare. For each of these areas, we have identified a number of initiatives and projects that illustrate how technology is changing the fields of medicine and healthcare. We have also identified a handful of case studies that make RRI tangible, highlighting how and why RRI is changing research and innovation practice.

A CHANGING LANDSCAPE FOR TECHNOLOGY AND HEALTH

The topic addressed by Sparks, ''technology shifts in health and medicine'', is deeply rooted in the big societal transformations that we are currently living through.

Technological developments take place today at such an accelerated pace that the institutions that were traditionally overseeing and advising about the impact of technology in society are finding it difficult to keep up with the complexity of the changing technological landscape. This creates, on the one hand, gaps in both understanding and regulating the use of new technology and, on the other hand, opportunities for new actors and innovators to address societal needs by swiftly adapting cutting edge research and disrupting traditional practices.

Responsible Research and Innovation (RRI) is the European Commission's approach to fit Europe with the capacity to manage the complexities and uncertainties that characterize contemporary research and innovation. Without the ambition of being exhaustive, the following trends introduce the context within which the activities of Sparks will be situated:

Systemic changes such as population ageing and increasing urban lifestyles are dramatically changing demands on healthcare systems across Europe. Currently, 18.5% of the population in Europe is older than 65 years; by 2050 this will grow to 28.1%. The number of people aged 80 and older will double, growing from the current 5.1% to 10.9%[1]. Combined with a diminishing workforce, these trends present unprecedented challenges to healthcare systems.

Consequently, the costs of current healthcare systems are growing at an unsustainable rate. There is a consensus among economists that most Western European countries today run healthcare systems that simply cannot be financed in the near future. The cost of healthcare is growing at a faster rate than is Gross Domestic Product (GDP), a situation that cannot be maintained. To remain viable, health systems need to change structurally. Simply put, short-term cost-cutting measures will not be enough.

The convergence of technologies, and in particular of nano, bio, information, and cognitive technologies (called ''the NBIC convergence'') is leading to unprecedented opportunities for innovation and investments, but also opens new terrains of ethical debate regarding the implications of these developments for human rights and human dignity. In addition to well-known ethical and social issues, such as safety, privacy, autonomy, responsibility, and informed consent, these

emerging technologies are raising new questions and dilemmas, such as: Should we have the right to enhance ourselves via technology? Do we have a right not be measured, analysed, and coached? How do we avoid being manipulated by human-like technologies? Which social tasks can we humanely delegate to machines? What knowledge, skills, and financial means do individuals need to use certain biomedical technologies appropriately in daily life? [2] We need new instruments, new modalities, new institutions and new venues for discussion to help answer these questions.

The use of medical devices and technology and the collection and analysis of biological data are no longer the purview of only medical research and healthcare delivery. For example, wristwatches and mobile phones now collect measurements about physical activity, nutrition, and other biological functions that are used to generate feedback and advice, bypassing a medical consultation. Genetic tests (and other testing and screening devices) can be easily bought on-line and performed at home. This leads to a "grey" area of activities that fall outside medical control and may or may not have regulatory approval but that can have profound consequences for the health and wellbeing of individuals and populations.

Current regulatory and ethical bodies (such as the Committee on Bioethics of the Council of Europe) are seemingly unprepared to deal with the rapid changes that are occurring across the medical field. Two major trends are disrupting the conventional approach to ethics: biology is becoming technological while technology is becoming biological. The result is an increasing diversity of practices and applications that escape current regulatory and ethical boundaries. Examples of developments where tensions are emerging between the constraints of current social, economic, and ethical boundaries and the possibilities enabled by technology include human and social enhancements, direct-to-consumer genetic testing, home use of medical devices, persuasive technologies, e-health, and mobile health.

Many technological developments that will eventually affect health and medicine will start and spread outside of the medical domain, thereby escaping the ethical norms, know-how, and regulatory framework of the medical field. The boundaries between medical practice, healthcare, entertainment, direct-to-consumer products, and human enhancements will become in- creasingly blurred and in some cases will even disap-

[1] http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing

[2] van Est, Rinie et al. (2014). From Bio to NBIC convergence – From Medical Practice to Daily Life. Report written for the Council of Europe, Committee on Bioethics, The Hague, Rathenau Instituut.

pear. A positive result will be increasing amounts of interdisciplinary research work across scientific domains, facilitated by and, in some cases, inspired by the lack of regulatory frameworks. A sign of this can already be seen in the living lab concept that is emerging as a test bed for aligning research with real users in real environments.

At the same time, regulatory wastelands will appear, areas where 'edgy' developments can proliferate virtually free from any constraint. These regulatory wastelands might function as social experiments for new types of emancipatory movements, allowing individuals and organized groups to both technologically and politically promote the use of biomedical technologies in the private domain. For instance, the 'hacking for health' movement is pushing the boundaries of synthetic biology and DIY (do-it-yourself) biotechnology, the 'quantified self' movement is exploring the possibilities of measuring and quantifying every aspect of human life, and the 'transhumanist' movement is advocating for technological enhancements to enhance the physical and mental performance of the human body and extend its senses.

RESPONSIBLE RESEARCH AND INNOVATION

RRI is a way of doing research and innovation where all involved parties pay great care to how today's decisions will affect the future. At its core, RRI is about envisioning the future and connecting research and innovation with that vision by governing our research and innovation activities towards ethically acceptable, socially desirable, and practically sustainable outcomes [3].

According to the European Commission, ''Responsible Research and Innovation (RRI) implies that societal actors (researchers, citizens, policy makers, business, third sector organisations, etc.) work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs, and expectations of society.'' [4]

RRI is informed by a set of values that the RRI Tools Project has identified as:

Democratic values regarding participation and power: this means that RRI is rooted in democratic concepts such as inclusion, diversity, openness, and transparency, and that all stakeholders are engaged in deliberative processes throughout the process of designing inclusive and sustainable research and innovation.

Social and moral values regarding care for the future of the planet and people: this means anticipating and responding to the needs of society, and delivering solutions that are socially desirable and sustainable.

Individual and institutional values of open-mindedness and receptiveness to change: this means that the actors and institutions involved in research and innovation are engaged in a continuous process to reflect on their role and practice, to learn from feedback, and to change accordingly.

The RRI Tools Project further provides a working definition of RRI that sums up the theoretical and empirical knowledge developed so far: ''Responsible Research and Innovation is a dynamic, iterative process by which all stakeholders involved in the Research and Innovation practice become mutually responsive and share responsibility regarding both outcomes and process requirements.''

The outcomes are:

- engaged publics
- responsible actors
- responsible institutions
- ethically acceptable research and innovation (R&I)
- sustainable R&I
- socially desirable R&I
- R&I that provides solutions to the grand challenges identified by the European Commission

The process requirements are:

- Diversity and inclusion
- Openness and transparency
- Anticipation and reflection
- Responsiveness and adaptive change

For an explanation of all these terms, please refer to the RRI-Tools Policy brief in the annex of this report[5].

Despite these definitions, RRI remains, however, a difficult concept to grasp. Many of the ingredients of RRI are rather vague and have different meanings to different people. Furthermore, RRI is a 'work in progress' that is being constructed and defined by the current research and innovation activities in Europe, including implementation projects such as Sparks.

To facilitate an understanding of RRI tailored to the scope of Sparks, we provide two approaches in the third section of this report that can be used to assess to what extent local projects and activities represent 'good RRI practice'.

[3] RRI Tools project – Deliverable 1.2 'Methodology for the collection and classification of RRI practices'

[4] http://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

[5] RRI Tools project – Deliverable 1.1 'Policy brief on the state of the art of RRI and a working definition of RRI'

HOW TO USE THIS REPORT

TIMEFRAME

The timeframe covered by the report is the next decade, with most of the case studies and examples taking place over the next 5 years. RRI is part of the Framework program Horizon 2020, so in principle it is an approach that will be implemented throughout Europe during the next 5 years leading up to 2020. The purpose of the report is to guide the reader on how to identify developments in healthcare related fields that are suitable topics for public discussion about the future direction of research and innovation. It is not intended to be a tool for foresight. For this reason, the report is not too speculative, but rather highlights current practices that give a tangible, realistic image of what RRI is about.

STRUCTURE

The report is divided into three sections, with the first section giving an overview on ***where to look*** for innovation in the medical and health fields. It describes six broad content areas where technological developments are shifting the way we think about healthcare. These areas are tightly interconnected and provide a 'canvas' on which research innovation can flourish.

The second section describes two approaches on ***how to understand and assess*** RRI in practice. One approach is rooted in transdisciplinary research and provides a straightforward set of quality indicators to assess to which extent a project embodies RRI components. The other approach is based on narrative structures and uses common storylines as tools to discuss RRI with the public.

The third section contains case studies exemplifying ***how to look*** for RRI in technological developments in healthcare. These case studies were selected to provide entry points to reflect about what it means to do research and innovation in a responsible way.

The feedback and observations gathered during the first Sparks workshop held in Amsterdam on 9 October 2015 were used to consolidate the report.

1. SECTION ONE

WHERE TO LOOK FOR TECHNOLOGY SHIFTS IN HEALTH AND MEDICINE

To organize and structure the content covered by this report, we have identified six broad content areas where technology is having a major impact on health and medicine. The process to define these areas included literature research, brainstorming sessions, and conversations with multiple stakeholders. Under each area we list current and future projects and initiatives that make tangible and visible the kinds of impacts that technological developments are creating in each domain.

The six content areas are tightly interconnected by what we call 'enablers', i.e. broad developments that enable cross-cutting research and innovation. These are major technological developments that are taking place outside the medical and health fields and that provide the overall frame of reference to understand the impacts of technology on health and medicine.

1.1 ENABLERS

Enabler 1

MOBILE TECHNOLOGIES

Mobile technologies are radically changing all aspects of healthcare, transforming delivery into a diffuse, granular, and pervasive system. They are shifting practices and instruments from being location-dependent (for example, practices and equipment based in hospitals or laboratories) to being location-independent. Tests and analyses can be done at the point of care, at home, and in fact anywhere and at anytime. Mobile technologies shift not only the location of the medical practice, but they change the medical practice as well. They empower citizens to perform tasks currently performed by medical and paramedical staff, such as diagnosis, counselling, and data journaling. They enable innovative approaches that challenge existing business models and established practices. Mobile technologies allow innovations to jumpstart: in Africa and Asia, for instance, mobile networks enable frugal innovation in services that are instrumental to providing healthcare, including financial and social services. In western countries, such innovations are constrained by the existing traditional infrastructure. *The trend towards personalized, predictive, preventive, and participatory health care can also be spotted in the research priorities mentioned in the work programme for 'health, demographic change and wellbeing' of the European Horizon 2020 Framework Programme. These developments depend on the ongoing miniaturization of devices and on making diagnostic tests increasingly available for use at 'the point of care', that is, at the bedside, in the consulting room, and also at home. As diagnostic devices become smaller, cheaper, faster, and smarter, they increasingly will allow people to test themselves and may even be incorporated within the body or brain. New types of information and communication technologies (ICT) also enable the arrival of tele- and mobile health care practices. ICT thus enables the monitoring and organizing of health care outside hospital settings.*[6]

Enabler 2

NANOTECHNOLOGIES

Nanotechnologies are going to revolutionize drug delivery and diagnostic devices, enabling a generational shift in how materials are used in healthcare. Thanks to nanotechnologies, we will be able to effectively work at the molecular and atomic level in the most complex biological structures, allowing detection of disease or malfunction at a level of detail that is unthinkable with current technologies. Nanotechnology offers new options for intervention inside the body based on nano-scale devices.

It supports new forms of large-scale data generation opening new possibilities for permanent monitoring of individual health states and for early diagnosis of disease. Alongside diagnosis, nanotechnologies will allow also a whole new approach to regenerative medicine, with the development of smart biomaterials (*Organs on a chip*, p. 24) that will blur the boundaries between living and artificial matter.

Enabler 3

DIGITAL TECHNOLOGIES

Digital technologies will become a digital layer fully integrated with physical reality. Digital information is becoming increasingly more location-dependent, and the availability of miniaturized and embedded devices to access information (wearables, implants, and in fact any object belonging to the internet of things) means that potentially every aspect of our lives can be outfitted with a digital layer of communication and information. The transformative effects of such pervasive availability of technology will be visible in new social norms, business values, and re-definitions of what ability and disability mean, for instance. Many materials with embedded communication capabilities will become programmable, so that environments and devices can be physically customized and respond to our presence and physical and emotional states. Developments in artificial intelligence and cognitive technology will break many of the current barriers between humans and machines and create a seamless integration bet-ween the digital and physical worlds.

Enabler 4

LIGHTWEIGHT INNOVATION

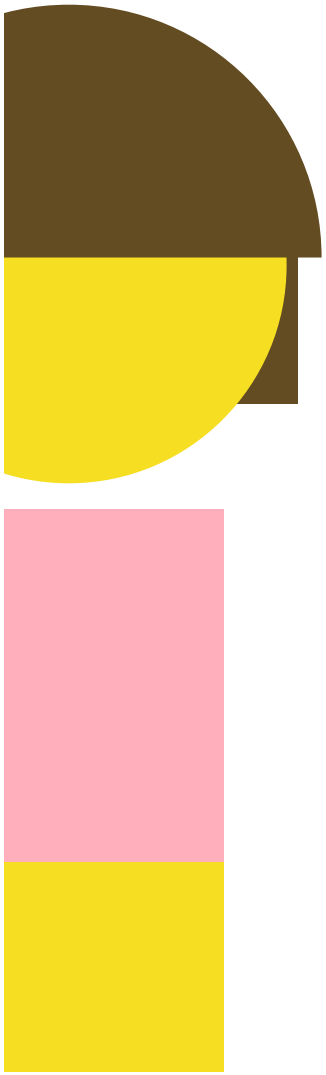
Lightweight innovation is a growing phenomenon where the process of innovation, usually dominated by industrial research and development (R&D), becomes widely accessible and distributed. The availability of technologies such as 3D printers and do-it-yourself biology instrumentation (such as handheld DNA sequencers) empowers a wide array of tinkerers and makers to do things that were prohibitive or impossible only a few years ago. These garage innovators are networked in a system that connects their specific competences, giving rise to an efficient system where the heavy, centralized and time-consuming traditional innovation process is replaced by a fast, distributed, and dynamic process. This approach to innovation will disrupt the current centralized model that has dominated R&D in the pharmaceutical industry for decades and will potentially unlock roadblocks to faster drug discovery and life-saving cures.

[6] van Est, Rinie et al. (2014). From Bio to NBIC convergence – From Medical Practice to Daily Life. Report written for the Council of Europe, Committee on Bioethics, The Hague, Rathenau Instituut.

1.2

SIX CONTENT AREAS

1.2.1 THE QUANTIFIED SELF



The convergence of nano, mobile, and digital technology makes it possible to accurately measure and record an infinite number of parameters related to the human body. From the number of steps one takes in a day to the quality of the cervical mucus, the quantified self allows us to monitor and get insight into so many aspects of our lives and it revolutionizes our concepts of self, identity, and human potential. Everyday objects such as mobile phones, watches, and glasses (and soon contact lenses) can be set to continuously track and monitor our movements, heartbeat, and many other functions of our bodies. Customized devices provide additional measurements and can be embedded in everyday clothing or even directly implanted in the body. The incredible level of detail obtained through tracking biometrical data can contribute to more timely, precise diagnosis and more efficient, customized therapies and medical procedures, but it will also challenge the cultural and social norms we currently hold. Curiosity and ingenuity are stretching the boundaries of how these devices are used. Sleep trackers, for instance, are being used by patients with Parkinson's Disease for disease management. The line between medical and non-medical use becomes impossible to draw.

WHAT TO LOOK FOR

The following are examples of initiatives and projects that exemplify the potential of the Quantified Self trend, as well as new questions that have arisen as a result of the unprecedented level of detail about the human body that technology is generating.

Wearables

Wearables are technologies that are worn on the body, usually as clothing or accessories (watches, earrings etc.), or as small implants or external devices. Aside from the multi-purpose watches and fitness trackers that are already very common today, some wearables address specific needs and provide solutions to problems that so far cannot be easily solved. For example, wearable thermometers in the ear can record changes in basal temperature with much better precision than before. This opens up new possibilities to better plan the timing of conception and increases the reliability of fertility awareness as a contraception method.

Profile Pruning

Because of the huge amount of personal data available and the existing sophisticated technologies for data mining, it is becoming increasingly affordable for companies and individuals to learn the health status and know about the risk profiles of other people. Therefore, some people will take action to conceal aspects of their digital identities by creating online alias personalities for specific activities, in order to disassociate such behaviours from their online profiles. However, this might be increasingly difficult: data analytics already infers health states from information never previously associated with health, such as driving habits or text messaging. For instance, depression can already be predicted with a high degree of reliability just by analysing mobile phone use patterns.

Biometric ID systems

Privacy and security depend on the correct identification of individuals. Traditional identification measures (photographs and fingerprints) are being complemented and substituted by much more precise systems such as retinal scans, face composition, and voice analysis. These technologies are not only more precise, but also more pervasive than before, and are challenging the concept of remaining anonymous. On one hand, they can be used easily without consent. On the other hand, they allow medical information to be transferred and accessed in a more effective way. This makes it paramount to ensure that only the persons who have the right to use the information can actually access it.

Embedded/ambient technology

Sensors are increasingly embedded into clothing, furniture, cars, walls, even trees and parks – in short, almost everywhere. They can monitor and record movement, physical activity, perspiration, location and proximity, while acting as real interfaces between the body, the physical environment, and the digital realm. For instance, Google is partnering with Levi's to produce fabrics that embed sensors and technology into everyday clothing. Technology is becoming embedded into everyday objects and can be completely invisible to the user and others. However, the data being captured may be invisible as well, meaning that we might not be aware of what information is being actually collected, how it is going to be used, and by whom.

WHAT TO ASK

How can we negotiate the extent to which technology becomes embedded in our bodies?

Are we still allowed to be anonymous? Are we redefining the concept of privacy?

How much do we know about the consequences of ubiquitous monitoring and awareness?

TO KNOW MORE

TED talk about the Quantified Self Movement
http://www.ted.com/talks/gary_wolf_the_quantified_self

The Quantified Self website
http://quantifiedself.com

Lifelogging – an exhibition at Science Gallery Dublin
https://dublin.sciencegallery.com/lifelogging

Google Jacquard
https://www.google.com/atap/project-jacquard/

European wearables producer Shimmer Sensing
http://www.shimmersensing.com

Fearing the quantified life – privacy, data and wearable devices
http://thenextweb.com/insider/2015/06/05/fearing-the-quantified-life-privacy-data-and-wearable-devices/

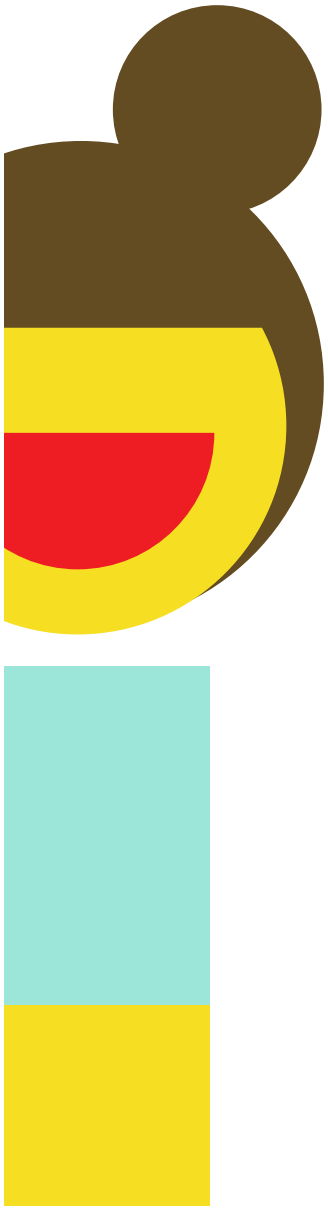
Using social media to detect chronic diseases and infectious disease outbreaks
http://vierds-app.nl/programma/sessie/2082/45
http://www.engadget.com/2015/08/30/princeton-harvard-rubella-kenya-phone-study/

Google and Novartis develop contact lens to help diabetes patients
http://www.forbes.com/sites/leoking/2014/07/15/google-smart-contact-lens-focuses-on-health-care-billions/

A wearable device to help become pregnant
http://www.fastcodesign.com/3049294/this-wearable-can-help-you-get-pregnant

The future of fitness – sport clothing with embedded sensors
http://www.liveathos.com/

1.2.2 *HAPPINESS*



In recent years the topic of 'happiness' has taken off as a holistic, comprehensive way to look at social progress, wellbeing and health. In 2012 the first World Happiness Report was published in support of the United Nations High Level Meeting on Happiness and Well-Being. The report is now a yearly publication with an important role in guiding policymaking worldwide[7]. Many countries have adopted happiness as an indicator of the nation's health and well-being. But happiness is not only measured: it can be constructed, controlled, and influenced. Recent research suggests that 50% of the determinants of happiness are genetic; 40% depend on our own state of mind, and 10% on contingent circumstances (such as social status, work, money, health, relationships etc.). While traditionally the genetic predisposition to happiness would have been considered outside of our control, it is becoming increasingly clear that genetic modifications within a lifespan are possible, influencing how we experience happiness[8]. Furthermore, advancements in neurosciences are beginning to provide a scientific basis to understanding what happens in our brain when we experience happiness. Four constituents have been identified so far: sustained positive emotion (experiencing short- and long-term pleasure and satisfaction); recovery from negative emotions; pro-social behaviour and generosity; and mind-wandering ('being distracted') versus mindfulness ('being focused'). Understanding these mechanisms of happiness (and probably others, still to be discovered) mean that with the appropriate technology it is possible to influence and directly affect how people experience happiness, modify their mood, and even change their behaviour.

WHAT TO LOOK FOR

The following areas represent developments where technology is being used to affect our state of happiness and blur the boundaries between health and entertainment, leisure and recreation, and lifestyle in general.

[7] http://worldhappiness.report

[8] http://www.theguardian.com/science/2015/aug/21/study-of-holocaust-survivors-finds-trauma-passed-on-to-childrens-genes

Body management

For several people, the awareness of one's body image is becoming the main determinant to make lifestyle choices, in particular those concerning eating and fitness. Eating and physical activity are therefore the main instruments to shape our bodies and our minds to achieve a preferred state. To satisfy this need (which depends in large part on the media and on social pressure) people often turn to non-medical products and methods that escape health regulations.

E-coaches

Almost any fitness tracking app now provides feedback in the form of badges, awards, goals, hints, and advice on how to "do better". Our data – the data we share with the app, at least – are analysed and the e-coach suggests the next step to take. The knowledge base available to an e-coach is often much broader than what a human coach can process and the resultant advice can be extremely precise. At the same time, are we in control of the information we share with the e-coach? Who decides on what priorities the e-coach works?

Gamification

Games are increasingly used as a method to achieve wellbeing. For example, there are games where players tackle depression, anxiety, insomnia, chronic pain and traumatic brain injury as if they were challenges in a game – and they solve them. The Mental Health Foundation of New Zealand used the evidenced-based principles about happiness mentioned in the previous page to create "The Wellbeing Game", an online game platform to increase personal awareness and mental health.

Persuasive media

Advertisements have been used for decades to influence people's attitudes and behaviours. Today the integration of behavioural sciences and information technology means that persuasive media can be highly personalized and trigger reactions and emotions at the level of the individual. Big data analysis, including sophisticated facial expression analysis, permits digging into and making sense of infinitesimal differences in how people react to media and then adapting the design of messages accordingly. Without consciously realizing it, we are driven into making choices that affect our health, wellbeing, and happiness.

WHAT TO ASK

Where lies the boundary between a medical and non-medical device?

Are persuasive media and transparency at opposite ends of the spectrum?

TO KNOW MORE

The wellbeing game
https://www.thewellbeinggame.org.nz

Jane McGonigal
http://janemcgonigal.com/

Persuasive media can be also low tech
http://www.bbc.com/news/blogs-news-from-elsewhere-31831559

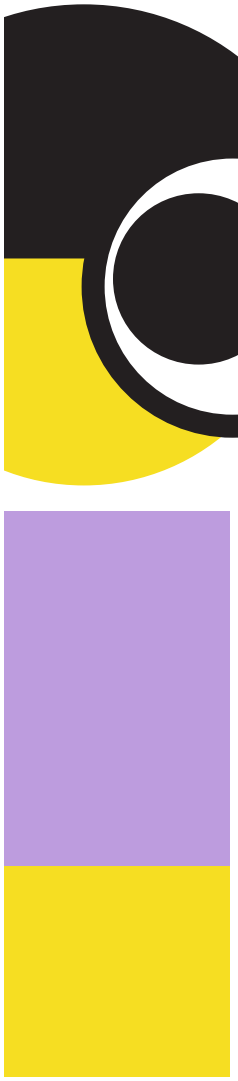
Facereader, a technology to measure emotions from facial expression.
http://www.noldus.com/human-behavior-research/products/facereader

Superbetter, a game to be stronger in difficult times
https://www.superbetter.com

How a text message persuades Swedes to donate more blood
http://time.com/3943272/blood-donation-sweden-text/

A dystopian prospect where every child can be screened for depression
http://gizmodo.com/testing-kids-for-potential-depression-seems-like-a-ba-1719856620

1.2.3 *BIO-REVOLUTION*



Biology can be technological and technology can be biological. This major shift has occurred as a result of advancements in our knowledge about the building blocks of life and by the broad availability of tools to modify, tweak, hack, construct, and re-construct biological materials. Synthetic biology, the interdisciplinary approach that creates new, synthetic forms of life, offers unprecedented possibilities to intervene in the human body to repair and restore organs and tissues, and even to add new functions and capabilities. But the possibility of actively manipulating life is not a revolution that we are witnessing at arm's length: biotechnology and synthetic biology are also becoming ubiquitous and affordable. Literally anybody can assemble the equipment necessary to sequence and synthetize DNA and to 'hack' living matter. Most of the software is free and open source, and the availability of 3D printers, coupled with programmable kits such as Arduino[9], opens the door to free and open source hardware as well. Miniaturization techniques make it possible to design integrated circuits that perform the same fluid analyses that traditionally are done in a full-size laboratory – but at fraction of the cost and using just a few millimetres or less of specimen. This shift in scale, known as LOC (lab-on-a-chip) revolutionises the whole concept of laboratory analysis. Results are produced much faster, analyses can be performed at any location, and mass tests can be done in parallel and at very low cost. The consequences of these new technologies are still to be understood. The possibility that such an 'on-demand', instantaneous, and inexpensive way of producing large amounts of biological data will structurally change our approach to health is very real.

WHAT TO LOOK FOR

The following are some of the most game-changing developments that give a sense of the tensions and opportunities around the biology revolution.

[9] https://www.arduino.cc

BioHacking

The hacker culture, born in the 1960s with the advent of computer programming, finds a natural evolution in contemporary biology, where life itself becomes programmable. The commonly accepted definition of hackers is that of people who enjoy the intellectual challenge of creatively overcoming and circumventing limitations in systems to achieve novel and clever outcomes. This 'playful excellence' stretches the limits of what is known and possible, and is one of the strongest forces behind the DIY-Bio movement. As with the hacking culture in computer programming, however, there are 'white hat hackers', who are driven by good intentions, and 'black hat hackers' who use their skills for evil and harmful purposes. BioHackers are an example in the growing field of non-institutionalised science and technology development.

h+

Technological advances in the biomedical field give rise also to new philosophical understandings of the human body. For example, transhumanists (known as h+) want to transform the human condition with the use of technologies that enhance the intellectual, physical, and psychological capacities of humans. Transhumanists base their beliefs on the parallel development of biology and computer science, leading towards the so-called ''singularity moment'' when computer-based artificial intelligence will be able to redesign itself and exceed human intellectual capacity and control.

Tissue engineering

Biological tissues can be grown in vitro and engineered to have specific mechanical and functional properties. Skin, bones, muscles, blood vessels, and other tissues can be 'harvested' in the laboratory and implanted in human beings. The next step is to grow entire organs which can replace defective ones or improve existing ones. Tissue engineering, which often uses stem cells, raises ethical questions about the origin of the living matter used to grow tissues. Merging biology with technology means also that biology, like technology, could potentially be patented. Currently, most regulatory systems are unprepared to deal with this area.

Organ-on-a-chip

Advances in microfluidics (the study of fluids in microscopic environments) permit the building of chips that perfectly simulate the workings of a human organ, in order to study the effects of drugs, bacteria, and other agents on the human body. At present, there are 'organs-on-a-chip' simulating lung, heart, kidney, liver, and arteries. Not only do these chips replace the need for in vivo and animal testing, they allow us to see new biological mechanisms and behaviours for the first time. These chips offer another revolutionary possibility: they can be designed to perfectly reproduce the unique composition of a specific person's organs, opening the way to highly personalized medicine. Furthermore, they can be networked to simulate the whole journey of a substance throughout the body. A new aerosol drug for instance can be tested as it enters the lungs, travels through the arteries, affects the

heart, is metabolised by the liver, and is finally excreted by the kidneys.

WHAT TO ASK

Are we comfortable with the concept of artificial life?

Can the human body be considered as a technology too?

Can you patent your own body?

How and by whom should decisions be made to advance the biorevolution further?

How and by whom should decisions be made to apply the technological advances of the biorevolution?

TO KNOW MORE

The Do-It-Yourself biologists (DIY-BIO) movement
http://diybio.org

Building research equipment with free, open-source hardware
http://www.sciencemag.org/content/337/6100/1303.full

organs on a chip
http://wyss.harvard.edu/viewpage/461/

The transhumanist movement
http://humanityplus.org

Detecting HIV with a mobile phone
http://www.wired.com/2015/02/hiv-diagnostic-tool/

DIY neuroscience kits
http://blog.ted.com/the-neuro-revolution-is-coming-greg-gages-neuroscience-kits-put-research-in-the-hands-of-the-curious/

1.2.4 *Global Health*

New innovations are revolutionizing the landscape of global health. Global collaboration is required to address major health problems that cross borders, whether between countries or transnationally through human air travel and transport of infectious agents or vectors. Examples of global health problems include zoonoses and antibiotic drug resistance, both of which are threatening human health.

A zoonosis is any disease or infection that is naturally transmissible from vertebrate animals to humans. Animals thus play an essential role in maintaining zoonotic infections in nature. Zoonoses may be bacterial, viral, or parasitic, or may involve unconventional agents. Major modern zoonoses include Ebola diseases and influenza. Mobile technologies combined with algorithmic forecasting will allow the development of more efficient early warning systems to contain disease spread and avert epidemics. Innovations in vaccine development and delivery are changing the range of tools at our disposal to fight zoonoses.

Drug resistance is a major problem threatening progress made in public health worldwide. Improper use of antibiotics has led to widespread resistant bacteria. In high-income countries, multi-drug resistant staphylococci (MRSA) are an alarming problem in hospitals and other care institutions. Worldwide, multi-drug and extensively resistant tuberculosis are undermining TB control measures. Portable and wearable technologies have the potential to improve drug adherence support and reduce the risk of mutations. Innovations in testing technologies, such as rapid high-throughput molecular assays will allow early detection of resistance mutations and enable new containment strategies.

Ubiquitous sensors, big data analyses, and the ability to prescreen for diseases will shift the intervention focus from constructing boundaries after-the-fact to anticipating the movement of people, animals, goods, and diseases[10].

[10] Institute for the Future (2010). 2020 Forecast: The Future of Science, Technology, and Well-being

WHAT TO LOOK FOR

These are some of the innovations that are changing our approach to global health problems:

Citizen science

The large scale involvement of lay people in research is leading to some major breakthroughs that were inconceivable just a few years ago. Personal computers and mobile phones are the main technological drivers that are transforming the role of citizen science in the medical field. In 2011 for instance, users of the citizen science platform Foldit discovered the structure of an enzyme critical for the replication of the HIV virus, a scientific problem that had been unsolved for 15 years. Thanks to open source technologies such as ResearchKit, mobile phones are becoming sophisticated instruments that can contribute an unprecedented amount of data and insights to research teams all over the world.

Big data

A major shift in how we approach global health problems is made possible by the capacity to store, retrieve, compute, and analyse huge quantities of data – the so called 'big data' revolution. For example, big data modelling permits complex simulations that can model the spread of infections in increasingly realistic ways. At the same time, the big data revolution means that information placed 'in the cloud' will persist there indefinitely. It is impossible to remove or withdraw information after it has been put in circulation.

Algorithmic forecasting

Algorithms are at the basis of predictive systems. Health insurance companies use them to predict the risk of insuring a new customer. They are used in medicine to help diagnose certain diseases. As the sheer amount of personal data increases, algorithms will increasingly be used to anticipate the risk of developing certain conditions. With growing computational power and sophistication, algorithms will facilitate the creation of 'just-in-time' and 'ahead-of-time' interventions for global health and personal care.

Patch vaccine

Advancements in nanotechnology have led to the development of a revolutionary system to administer vaccines, based on an array of thousands of vaccine-coated microneedles. These needles, about 20.000 on a 1 cm^2 surface, release the vaccine directly to the immune cells under the skin. Pre-clinical tests show that this technology reduces the amount of vaccine necessary to achieve efficacy by 100-fold while simultaneously amplifying vaccine efficacy. Furthermore, the patch vaccine does not need refrigeration in order to maintain a cold chain from production to administration, making it suitable for use in all kinds of environments.

WHAT TO ASK

Do we need new ethical frameworks to address global health problems?

What is the balance between sharing personal information to make better forecasting systems and respecting individual privacy?

What constitutes a 'responsible' use of antibiotics?

TO KNOW MORE

Lifesaving innovations with great promise to transform global health by 2030
http://ic2030.org/2015/09/sdgs-rd/

Patch vaccine
http://www.ted.com/talks/mark_kendall_demo_a_needle_free_vaccine_patch_that_s_safer_and_way_cheaper

Phone call use can predict infectious disease outbreaks
http://www.engadget.com/2015/08/30/princeton-harvard-rubella-kenya-phone-study/

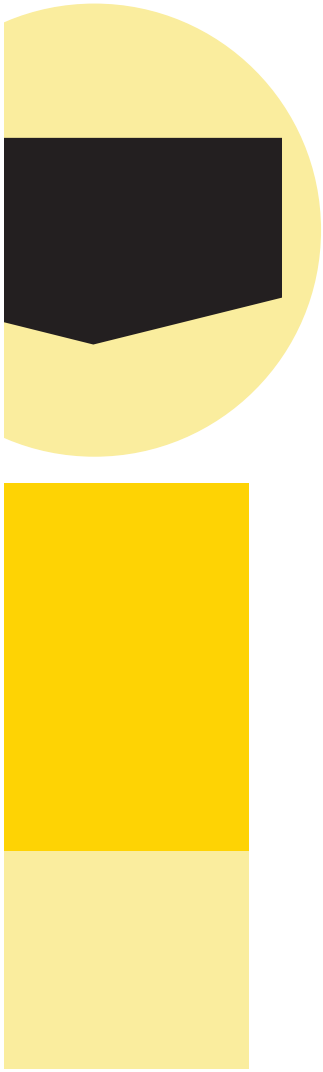
An inexpensive chip to contain the spread of Ebola
http://www.cnet.com/news/new-technology-could-help-contain-spread-of-ebola

An algorithm can cure us from cancer
http://www.salon.com/2015/10/10/an_algorithm_might_save_your_life_how_the_amazon_and_netflix_method_might_someday_cure_cancer/

Foldit: Solve Puzzles for Science
https://fold.it

ResearchKit
https://www.apple.com/researchkit/

1.2.5 *ROBOTICS*



The robots are coming! We have seen and read countless science fiction stories about the advent of robots, but it seems now that it is for real: the robots are finally coming among humankind. To some extent, they have been among us for quite some time now: automation in the manufacturing sector, for instance, is nothing new. Today, however, the integration of artificial intelligence, miniaturization, and 3D printing means that mechanical droids act as full interfaces between the physical world and the digital realm. It means also that we, humans, increasingly need to interact directly with robots. Robots do not stay anymore behind the scenes but take front stage and become useful companions and assistants for healthcare and medical practices. In fact, robots can be considered programmable matter, elements of reality which can be programmed and directed to perform certain tasks. This 'physical amplification' of data means also that mathematical models can be brought to life – not only metaphorically – by creating bionic extensions to the human body. Currently these are used to replace bone structures and lost limbs. Such implants can be highly personalized and can interface with the neurological system in order to fully integrate with the human body. Increasingly, bionics will add more structures to the human body, such as a 'second skin' or external support structures ('exoskeletons').

WHAT TO LOOK FOR

What are the current areas where robotics and healthcare meet?

3D printing

3D printing has brought the possibility of manufacturing a wide range of materials on-demand. It is opening up unprecedented possibilities in health and medicine. The number of 3D printers is growing exponentially; they are becoming also much faster and economical. Today it is possible to print with conductive materials and embedded chips, so that what is manufactured can be directly connected and networked. Customized, on-demand replacement parts and body enhancements are becoming possible, ubiquitous, and affordable.

Embedded/ambient sensors

Home automation is a research and manufacturing field that studies and produces sensors and devices for home use. Intelligent thermostats and lighting systems that can be connected and remotely controlled are already successful consumer products. Embedded sensors can be used to gather a wide range of information about people's movements, temperature, sounds, and the presence of micro organisms and other particulates. Analysis of these data can be used to alert regarding dangers and to devise strategies for a healthier life. But collecting this data also represents a threat to privacy and can encounter public opposition.

Drones

Drones are small flying devices that can be remotely operated. They can be used to monitor areas that are not accessible to humans; they can transport and retrieve deliveries of any kind. For example, they can be used to deliver medical equipment and drugs to areas that cannot be reached by other means because of lack of infrastructure or the impossibility of using existing infrastructure. Miniaturization will soon make drones incredibly small – to the point of making it really possible to remotely be 'a fly on the wall'.

Second skin and exoskeletons

The skin is our interface with the external world. There are a number of technologies today to develop and grow artificial skin, but the real challenge remains a skin, like our own, that can sense touch and interface with the nervous system. Recently a new layered technology has been developed that can connect skin to neurons in the brain to recreate the feeling of touch, opening the way to fully functional bionic prosthetics. An exoskeleton is a hard structure that supports the body and facilitates movements. It is like a skeleton, but it resides outside the body. It can assist when muscle power is limited or when a person cannot coordinate movements well, for instance because of neurological problems.

Robot nurses

Robots are being tested as companions for elderly people, for people suffering from autism, and in other situations where human care is being given now. The purpose of robots is not to substitute for human care but to provide a level of care that cannot be achieved by humans. It may sound ambitious, but there are already several situations in which robots are better accepted than humans. For a more detailed description of this field, see the case study *Alice*, on p. 52.

WHAT TO ASK

How big is the step from therapeutic robotics to human enhancements?

Are we really moving towards a singularity moment?

TO KNOW MORE

Touch sensitive skin
http://arstechnica.com/science/2015/10/luke-skywalkers-prosthetic-arm-inspires-artificial-skin/

Using drones to deliver medicines
http://gizmodo.com/abortion-drone-is-the-best-drone-1713388194

Hackers can enter into medical devices as well
http://time.com/3983847/hackers-medical-devices/

Low cost prosthesis
https://waag.org/en/project/low-cost-prosthesis

LEGO prosthetic arms for children
http://www.theguardian.com/artanddesign/architecture-design-blog/2015/jul/22/lego-prosthetic-arm-that-kids-can-hack-themselves

3D printing prosthetics
http://techcrunch.com/2015/08/20/watch-a-girl-named-isabella-unpack-a-new-3-d-printed-arm/

1.2.6 *OUR DEEPER SELVES*

Modification of the nervous system from the outside (for instance, by trepanation and electroshock) has been practiced for centuries, or even millennia, raising all sorts of ethical questions and social controversy. Today we can rewire the brain just as we modify the body with a tattoo or a piercing. We can use a ray of light to directly change the neuronal network inside the brain. Combining advancements in genetics and neurology, optogenetics is a breakthrough method to modify living cells – mostly neurons – using light, after the cells have been modified to be light sensitive. At the same time, less invasive technologies are becoming mainstream. Virtual reality and augmented reality are already being used, for example, to rewire the brain after a stroke and also in autism. Consumer products are currently being sold that 'zap' the brain and induce states of relaxation or alertness. Eventually this technology may become as pervasive as smartphones are now. What some people believe to be ethically questionable techniques are becoming lifestyle 'enhancements'. The widespread availability of technology is re-defining the ethical boundaries to which we are accustomed.

WHAT TO LOOK FOR

These are a few examples of how technology is improving our understanding of the deepest systems in our body and how it allows an unprecedented level of manipulation of these systems.

Virtual Reality (VR)

The quality of virtual reality technology is already so high that it is currently used in the training of surgeons and physicians as an addition to practicing on actual bodies. This offers unique opportunities to expand and renew the training of the medical profession. For instance, it makes possible to use an unlimited diversity of cases on which the medical students can practice. The quality and approximation of the models remains however an open issue. On the consumer side, new affordable devices provide immersive experiences that 'trick' the brain into believing, with all senses, to be in another environment. Game based and physical rehabilitation is a growing area of use. The possibility to use VR to reprocess memories and 'live' alternative scenarios makes it also a powerful tool for psychotherapy.

Transcranial direct current stimulation (tDCS)

This technique consists on stimulating different areas of the brain with a low electrical current. Various trials have shown that tDCS can be effective in increasing attention span, memory, and mathematical ability, as well as in treating depression. There is growing evidence that it has useful therapeutic applications for patients with Parkinson's disease or Alzheimer's disease. Because it is generally considered a safe technique, consumer products are now available that use tDCS to induce states of relaxation or improved focus. This controversial use of medical procedures in non-medical settings is at the moment completely unregulated.

Epigenetics

Environmental factors, age, and stress can affect the way genes are switched on or off, causing differences in the expression of a person's traits because of how the DNA is read, rather than because of actual changes in the DNA. This means that external factors can change how our body reads the genetic code, and these changes can be inherited by the next generation. Epigenetics is the study of how this mechanism works, including the hereditary aspect of it. Epigenetics can offer revolutionary new ways to understand cancer and mental disorders, for example. But it also shows that it is possible to modify the hereditary traits of a person using external factors.

Artificial intelligence

Developments in artificial intelligence proceed so fast that some scholars believe there will be soon a moment where computational capabilities exceed human intellectual capacity. At that stage, artificial machines can take over the task of designing themselves and evolve of their own. This is the so-called 'singularity moment'. While it might sound very futuristic, it is believed this will happen within the next 10-20 years. In the meantime, artificial intelligence is a backbone of big data analysis. It can provide sophisticated decision-making systems that rely on patients' data. This power of analysis leads to new medical knowledge, but also to resistance within the medical field because of the major shift that it represents in the medical practice. There is no threat yet that machines will substitute for physicians. However, the training that physicians

receive might need an overhaul to make sense of and usefully use the cognitive power provided by artificial intelligence.

WHAT TO ASK

Is it appropriate to regulate artificial intelligence, or will it regulate us?

Can we consider consumer tDCS devices as being on the same level as recreational drugs?

To what extent is virtual reality simply becoming the reality and how concerned do we need to be?

TO KNOW MORE

How Virtual Reality is impacting medicine and healthcare
http://www.techrepublic.com/article/10-ways-virtual-reality-is-revolutionizing-medicine-and-healthcare/

On affecting evolution by changing our genetic system
http://www.ted.com/talks/harvey_fineberg_are_we_ready_for_neo_evolution

An evening with the Consciousness Hackers
http://www.newyorker.com/business/currency/an-evening-with-the-consciousness-hackers

The ethical dilemmas of human genetic modifications
http://qz.com/441423/why-china-wont-listen-to-western-scientists-about-genetically-modifying-the-human-embryo

Zapping our brains
http://www.huffingtonpost.com/entry/brain-stimulation-thync_55b15ed7e4b0224d88319a6d

A cognitive computer systems to help solve health problems
http://www.engadget.com/2015/04/13/ibm-watson-health-cloud/

Artificial intelligence can tell what kind of person you are from your emails
http://mashable.com/2015/07/27/work-email-ibm-watson/

Neural implants give hope to paralyzed ALS patients
http://spectrum.ieee.org/tech-talk/biomedical/bionics/neural-implant-enables-paralyzed-als-patient-to-type-6-words-per-minute

Many of the signals identified under each area are in fact multidisciplinary. For instance, at the intersection between Our Deeper Selves, Robotics, and Bio-Revolution we find human enhancements (both chemical and physical).

2. SECTION TWO

HOW TO UNDERSTAND AND ASSESS RRI

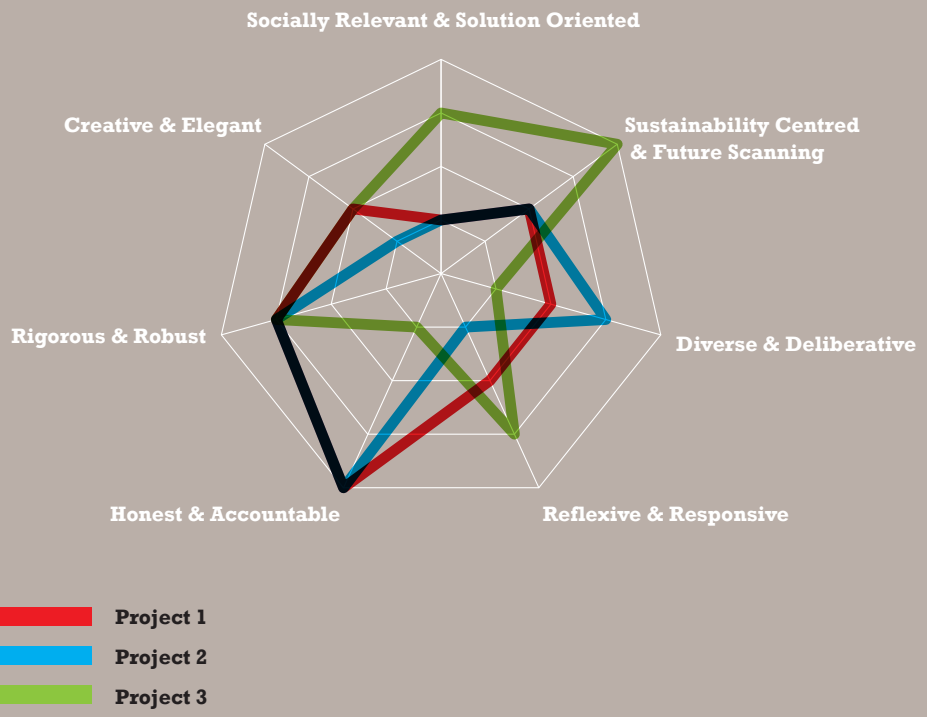
In addition to the definitions of RRI described in the introduction of this report, we propose here two additional methods that can be used to determine the RRI aspects of a project or an activity. These methods can be used to identify the content for the exhibition, as well as the case studies and examples for local events. It is very important to acknowledge up front that these methods are not meant to be a clear-cut black and white procedure to tell if RRI is present or not in a certain activity. RRI is a complex and dynamic understanding; reducing it to a series of rules and closed options would completely undermine its meaning. We recommend therefore that local organizers use the resources developed by the RRI Tools project to help them to identify examples and best practices of RRI. However, we realize that local organizers might find it useful to have a perspective on RRI that is specifically tailored to the needs of the Sparks project.

2.1 A RUBRIC OF QUALITY CRITERIA AND INDICATORS

The first method is based on the research by Fern Wickson and Anna Carew [11]. They propose a set of quality criteria and indicators (a 'rubric') to identify to what extent a certain research or innovation activity responds to the 'qualities' of RRI. The authors developed the quality criteria on the basis of the following seven characteristics of a research and innovation activity:

Socially relevant and solution oriented
Sustainability centred and future scanning
Diverse and deliberative
Reflexive and responsive
Rigorous and robust
Creative and elegant
Honest and accountable

The appropriateness of this approach for Sparks resides in the fact that the authors provide a set of standard rules to assess if an activity is an excellent, great, good, or routine (that is, not so good) implementation of RRI. This rubric, which is similar to that used in higher education to assess and provide feedback on student work, can be used to identify excellence and possible pitfalls of research and innovation projects in an objective way. Furthermore, it can also be used to compare different projects on the basis of quality criteria. To visualize differences across projects, a spider chart can be used to plot where each project stands on the seven RRI characteristics.



This is a straightforward and relatively easy method to help determine where a research project falls on a scale from ‘routine’ to ‘excellent’ in terms of the seven RRI criteria. When selecting the content for the exhibition and for local events, Sparks local organizers can use the criteria and indicators in this table to guide their choices. Note that the criteria in this table are slightly different from those used in the definition of RRI developed by the RRI Tools initiative, as reported in the introduction section of this report. However, there is a very large overlap between the two; in fact, the differences represent a good example of the complex and dynamic nature of the concept of RRI.

[11] Wickson, F., & Carew, A. (2014). Quality criteria and indicators for responsible research & innovation: Learning from transdisciplinarity. Journal of Responsible Innovation, 1:3, 254-273. doi:10.1080/23299460.2014.963004

CRITERIA

to which extent is the activity …	Exemplary: if it …
Socially relevant and solution oriented **What type of problem does it address?** **What type of solution does it aim to achieve?**	• addresses a grand social challenge.[12] • analyses objectives and processes on an on-going basis to favour delivery of 'wicked solutions' that solve multiple challenges simultaneously[13].
Sustainability centred and future scanning **How does it identify potential risks and benefits; does it consider social, economic, and environmental sustainability?**	• includes formal processes of future forecasting at various points throughout the research and innovation process. • identifies and assesses risks and benefits; generates a range of positive and negative future scenarios for social, environmental, and economic sustainability; • has clear methods for embedding responses to risk/benefit assessments and to possible future scenarios in project development.
Diverse and Deliberative **How interdisciplinary is the activity?** **How and when does it involve the stakeholders?**	• openly and actively seeks ongoing critical input, feedback, and 'feedforward' from a range of stakeholders. • encourages and rewards transformative mutual learning. • employs an evolving integrative method and consciously employs a transdisciplinary process.

Great: if it …	Good: if it …	Routine: if it …
• addresses a significant social need. • analyses objectives and processes on an on-going basis to maintain a focus on delivering a successful solution.	• focuses on a marginal or self-defined problem. • employs processes aimed at generating insights toward a solution or a partial solution.	• pursues a purely personal interest. • possibly delivers only decontextualized knowledge or decontextualized new problems.
• includes future forecasting activities at some point during the research and innovation process. • does some attempt to integrate an assessment of the risks and benefits and to consider social, environmental, and economic sustainability. • identifies points and possibilities for adaptation of processes to respond to risk/benefit assessment activities and to future scanning findings of importance to sustainability.	• includes informal attempts to forecast the future at limited points in the project. • includes a consideration of some associated risks and benefits and of some social, economic, and/or environmental sustainability issues. • gives little indication of how the research and innovation process may adapt and respond to identified possible risks or to future scanning findings of importance to sustainability.	• does a singular optimistic prognosis for future outcomes of the project, with no clear effort to identify risks or survey alternative future scenarios.
• invites, incorporates, and integrates stakeholder views at various points along the research and innovation process. • actively seeks dialogue and interaction with stakeholders; is open to mutual learning. • encompasses a wide range of methods and adopts an interdisciplinary process.	• takes limited steps to engage stakeholders in various stages of the research and innovation process. • has a tendency toward one-way forms of communication with stakeholders but is open to some interaction. • involves some level of methodological diversity and multidisciplinary practice.	• communicates with stakeholders only toward the end of the research and innovation process. • uses one-way communication approaches and is defensive in the face of countervailing views or stakeholder questions. • is mono-methodological and mono-disciplinary.

[12] http://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges

[13] 'Wicked solutions' try to solve 'wicked problems': https://en.wikipedia.org/wiki/Wicked_problem

	Exemplary: if it …
Reflective and Responsive **Does the activity recognize the preconditions that exist in the context and the group where it is located? Does it explore the underlying values, assumptions, and choices made? Is it open to critical scrutiny and does it show ability to change after internal reflective practice and external feedback?**	• identifies clearly and explicitly institutional and contextual limitations and demonstrates a structured effort to acknowledge and improve upon these conditions. • performs a structured, purposeful periodic analytical review of underlying values, assumptions, and choices. • actively seeks critical feedback from a wide variety of sources and actors. • shows evidence of potential to adapt at various points in response to internal reflective practice and external review/input/feedback.
Rigorous and Robust **How does it investigate the problem? Which aspects of the problem does it consider? Are the results repeatable across different actors and settings? Are the results reliable under real-world conditions?**	• includes a comprehensive investigation of all aspects of the problem and the interconnections between them. • generates results that are repeatable by a variety of different actors operating across a range of relevant conditions. • delivers outcomes that work reliably under real-world conditions.

Great: if it …	Good: if it …	Routine: if it …
• makes an explicit effort to identify institutional and contextual limitations and demonstrates awareness of their significance for practice. • makes occasional use of a structured process for reflecting on underlying values, assumptions, and choices. • actively seeks critical feedback from select sources and actors. • clearly indicates a capacity to adapt in response to reflective practice and external feedback.	• shows some indication of awareness concerning institutional and contextual limitations. • employs an informal, one-off or ad hoc process for considering underlying values, assumptions, and choices. • accepts critical feedback when offered. • is willing to accept change in response to internal reflective practice or external review and critique.	• gives no explicit consideration or recognition to institutional and contextual limitations. • has no process for facilitating reflective practice. • does not actively seek critiques. • shows no evidence of potential for change in response to criticism and unsolicited feedback.
• considers multiple dimensions of the problem and their interrelations. • generates results that are repeatable by the same actors operating under a range of relevant conditions. • delivers outcomes with demonstrated functionality under real-world conditions.	• investigates several dimensions of the problem although not necessarily their interrelations. • generates results that are repeatable by the same actors operating under similar conditions. • delivers outcomes that remain untested under real-world conditions.	• has a narrow focus on one element or aspect of a problem. • generates results that can not be replicated. • delivers outcomes that can not be reliably applied and replicated in real-world contexts.

	Exemplary: if it …
Creative and Elegant **To which extent is the approach innovative and daring?** **To which extent does it value efficiency and beauty in its approach?**	• reframes the problem in innovative directions, with new ideas being pursued through appropriate methods. • carefully considers resource requirements and allocates resources efficiently to achieve maximum utility and impact. • gives consideration to the aesthetics of preconditions, processes, and products.
Honest and Accountable **How does the activity identify uncertainties and limitations?** **What are the lines of delegation and ownership?** **Does it comply with research ethics and government requirements?** **What are the policies on open access and information sharing?** **Does it promote ownership of positive and negative outcomes?**	• identifies in a transparent way the range of uncertainties and limitations that may be relevant for various stakeholders. • communicates openly about lines of delegation and ownership that are able to respond to process dynamics and contextual change. • Documents and monitors actively on an on-going basis its compliance with highest-level governance requirements, research ethics, and voluntary codes of conduct. • consistently implements open access information policies. • demonstrates accountability for both potential positive and negative impacts.

Great: if it …	**Good: if it …**	**Routine: if it …**
• develops new methods according to new ideas within an established problem framing. • explicitly justifies the use of resources. • gives consideration to the aesthetics of preconditions and products.	• pursues new ideas through established methods within an accepted problem framing. • employs considerable resources in an inefficient way. • gives consideration to the aesthetics of envisaged products.	• frames the problem, the ideas, and the methods within established paradigms. • dedicates extensive resources (e.g. time, money, personnel, etc.) to work that has minimal significance or potential impact. • gives no consideration to the aesthetics of operating preconditions, research and innovation processes, or envisaged products.
• identifies uncertainties and limitations that are deemed to be significant by those involved. • has established lines of delegation and ownership. • complies with governance requirements and research ethics with evidence of active monitoring throughout. • favours open access information policies. • promotes accountability for potential positive and negative impacts.	• provides some statement indicating awareness of uncertainties and limitations. • indicates potential lines of delegation and ownership. • complies with minimum standards of governance requirements and research ethics. • occasionally employs open access information policies. • promotes accountability for positive impacts and some negative impacts.	• is not transparent concerning limitations and uncertainties. • does not allow tracing of the ownership of components. • does not specifically acknowledge standards concerning governance requirements or research ethics. • has no demonstrated commitment to open access information policies. • promotes accountability only for positive outcomes.

2.2 THE NARRATIVE APPROACH

Another approach that can be used to identify and prepare the content for Sparks is the narrative approach, inspired by the research of Macnaghten, Davies, and Kearnes[14]. Rather than being a methodology to identify the RRI components of a certain activity, the narrative approach helps develop different ways of presenting content to the public, based on three recurring and interplaying perspectives that become evident when the public is confronted with emerging technologies. Macnaghten, Davies, and Kearnes have found that public attitudes and concerns about new and emerging technologies tend to follow three main storylines or narratives. These narratives cut across national and cultural boundaries and are considered by some people to belong to the shared heritage of Europe.

THE THREE NARRATIVES

1. A dominant master narrative of scientific breakthroughs linked to social progress. This optimistic narrative includes those stories with a strong influence from the Enlightenment, where knowledge ultimately leads to progress and to a better situation for all. Under this narrative we find stories where science and technology reduce disease and illness, increase health, strengthen social cohesion, etc. These narratives are usually associated with breakthroughs achieved by science: for example, the eradication of smallpox or the discovery of antibiotics. Even when a group effort is acknowledged, these narratives tend to highlight the work of one scientist or innovator who had the genius to make a discovery and create positive impact. In the case of contemporary science, these stories project a future where science and technology allow humankind to live better lives, to reduce social and economic disparities, to take care of the environment, etc.

2. An 'ancient' counter narrative where transgression of natural orders and boundaries (hubris) leads to ills and harms (nemesis). Under this narrative we find stories belonging to the 'be careful what you wish for' tradition, where enthusiasm about achieving high expectations (sometimes unrealistically high, or acting too fast and without proper reflection) is curbed by unpredictable and unforeseen disasters and consequences. Another common storyline is the 'Pandora's box' narrative that provides an explanation for all the evils and miseries of the world. In this case, the accent is on thoughtlessness associated with our incapacity to anticipate future problems and on our incapacity to turn back once the technology is 'out there'. A third example is the

'messing with Nature' narrative, where humans are not supposed to cross the moral and ethical boundaries set by nature, which itself is considered sacred. In this case, the technoscientific endeavour is considered 'irresponsible' because it disrupts natural orders and boundaries. It is not negative per se, rather what is negative are the ideas of messing, tinkering, and interfering with nature, and of extending control over the natural world.

3. A 'modern' counter narrative where publics are exploited and alienated through technology. One example is the 'kept in the dark' storyline, where people are powerless and unaware of technological developments, either because these are in the hands of an elite (governments, corporations, media) or because technology itself develops in ways that are uncontrollable by society and politics. In this narrative, people associate technology with secrecy, they experience a lack of agency, and they feel compelled to trust 'expert systems' (governments, scientists, media) that they deem necessary to govern technology but over whose conduct they feel completely powerless. Another narrative is 'the rich get richer' storyline about social inequality and injustice. Even purportedly inclusive technological developments will end up meeting the logic of a neo-liberal political economy, with the rich (big business, the wealthy, and powerful elites) benefitting, while the poor are further marginalized. It incorporates powerful notions of morality, fairness, and justice. Here belongs also the pessimistic narrative that new developments in technology will not materialize soon enough for most citizens, given how long it took in the past to gain widespread access to innovation.

How can this approach be used in Sparks? These narratives help to develop storylines for the exhibition and for the events in which the public will recognize their own aspirations, concerns, and expectations. While these narratives are not exhaustive of all the possible narratives that the public can employ, they do appear repeatedly and consistently and they have emerged from the analysis of several dialogue initiatives and numerous focus groups. These narratives can also be found in the recent Eurobarometer qualitative study on public opinions on future innovation, science, and technology published by the European Commission[15]. In this study, citizens from 16 European countries were asked to elaborate on their opinions about current developments in science and technology affecting health and healthcare and to comment on a scenario describing a possible future. Across countries, citizens' responses mapped on to these three major narratives.

[14] Macnaghten, P., Davies, S. R., & Kearnes, M. (2015). Understanding Public Responses to Emerging Technologies: A Narrative Approach. Journal of Environmental Policy & Planning. doi: 10.1080/1523908X.2015.1053110

[15] http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/Survey/getSurveyDetail/instruments/QUALITATIVE/surveyKy/2007

Citizens expect that innovations in healthcare will lead to improved quality of life, increased life expectancy, and more control over one's health. These are examples of the first narrative, a storyline where science improves human condition. At the same time, there are negative expectations about the social and health consequences of longevity, the ethics of certain treatments (such as genetic manipulation and cloning), and the likelihood that rising costs will lead to unequal access. These concerns are examples of the two counter narratives. For example, concerns about moral and ethical issues of keeping people alive for longer and about the ethical acceptability of treatments such as in vitro fertilisation (IVF) and cloning belong to the narrative about transgression of natural orders. Concerns about the negative consequences of rising healthcare costs relate to the modern narrative of social injustice.

Additional narratives that emerged during the Sparks kick-off meeting on 7-8 July 2015 were:

- A narrative of 'knowledge is power' that highlights how access to knowledge and the opportunity to contribute and shape knowledge are powerful instruments for all members of society.

- A narrative of people daring to do unconventional things and to break conventional approaches to research and innovation. This 'daring' attitude needs to be balanced with inclusiveness, reflection, and anticipation.

- A narrative of sharing risks across stakeholders to achieve better social justice and equality.

The three main narratives and these additional narratives can be used as a 'canvas' or a framework to prepare and present the local content of the exhibition and the events from different perspectives and with different voices, and to raise the importance of RRI in addressing the consequences of each narrative. Taken together, these narratives present the points of view, concerns, and ambitions of multiple stakeholders. They are a good starting point to open up the discussion about the implications of tacit assumptions about nature, science, technology, and progress that may be at odds with wider public sentiment.

Using the three main narratives, the questions we want to answer are:

Narrative of scientific breakthroughs linked to social progress (triumph of pure knowledge – Enlightenment):

- Is there idealism behind it?
- How is knowledge produced?
- Who identifies the problem?
- Who leads the process?
- Who owns the process?

Narrative where transgression of natural orders and boundaries (hubris) leads to ills and harms (nemesis):

- What are the obstacles, accelerators and facilitators?
- Where is the unknown?
- What is defeat?
- Who leads the process?
- Who owns the process?

Narrative where publics are exploited and alienated through technology; where technology and wellbeing are unequally distributed:

- Who pays?
- It is possible, but is it necessary?
- Does it increase social equity?
- Who leads the process?
- Who owns the process?

Creating narratives to present and discuss new technologies will help to question some of the assumptions held by the stakeholders involved in RRI, including the local organizers and public. It will also facilitate collaborations and interdisciplinary work with the humanities, arts, social sciences, philosophy, and other domains.

3. SECTION THREE

HOW TO LOOK FOR RRI

Arguably, there is no single research or innovation activity that can be considered exemplary with respect to all aspects of RRI[16]. This is due in part because RRI is a relatively new concept and it will take some time before it is fully integrated in the research and innovation practice. In addition, definitions of RRI are still fluid and the criteria that can be used to identify examples of RRI are still in development. It is important to remember therefore that a suitable case study for Sparks does not mean that it is a perfect example of RRI, but rather a good starting point to engage the public on RRI.

The criteria to identify a good case study are broader than the aspects of RRI alone. They include considerations of how appealing the topic will be for a museum/science centre audience, the amount of information and media available on the topic and about the research process, and how widespread and relevant is the research on a specific subject in Europe.

To identify the case studies, we drew on discussions held at the first Sparks workshop on 9 October 2015. A key concept that emerged was that although it is common to identify 'research' as a process that is structured and formalized in institutions (such as universities, enterprises, and research centres), in fact 'research' often starts and develops outside of these institutions. It is a process that stems from informal and unstructured conversations among a large number of individuals who are not yet considered 'stakeholders' in the formal sense because what is at stake is not yet defined. It is important therefore to capture how RRI applies to these non-traditional spaces and processes, which are becoming increasingly more important with the widespread availability of technology.
We present therefore two kinds of case studies: a specific research pro-

[16] RRI Tools project – Deliverable 1.4 'A catalogue of good RRI practices'

ject, grounded in interdisciplinary academic practice, which is exemplary of how RRI practices can be identified from available documentation; and an assemblage of emerging themes that, even though still fragmented in terms of research practice, provide examples of how RRI can be discussed with the public at this very early stage in the research and innovation process.

Case Study 1

ALICE THE CARE DROID

Demographic change – and in particular an ageing population – is one of the major challenges in Europe. In Amsterdam, for instance, it is expected that by 2024 there will be four times as many people older than 80 years as there are now. In Europe, on average, the number of people aged 80 and older is going to double between now and 2050. As a consequence, the need to provide care to elderly people will increase dramatically. However, there are substantial uncertainties that this need can be met with the conventional approach to care and assistance, which is currently based on the human workforce. The problem is not only the availability of personnel to cope with the increasing number of elderly people, but also the different kinds of care that will be needed in the future. Because of the general advancements in medicine, elderly people will live healthier lives and the care they need will be less medical and more social. It will be important for elderly people to remain socially active as part of healthy ageing.
In 2011 a research consortium comprising Dutch universities, robotics manufacturers, and designers started to investigate[17] what it would take to develop robots that can provide necessary care to elderly people.

Initially, the research approach was to design robots that are emotionally intelligent and human-oriented: in short, robots that emulate as much as possible the standard of care that humans can provide. This involved researching and developing artificial intelligence systems that can provide robots with emotional, moral, and ethical reasoning. The research team prepared a prototype of a care droid and they tested it with a group of elderly ladies, to observe how they react and behave with a robot that can listen and talk. In this test, a technician was operating the robot

behind the scenes simulating the feedback that fully functional artificial intelligence systems would provide. A documentary movie was also produced about the whole test. The documentary was shown at various film festivals and on Dutch national TV. It won the first prize for science communication media from the Dutch Research Council and the first prize at the international science film festival Pariscience.

This test, together with feedback from focus groups, observations, and conversations, radically shifted the design approach and the goals of the whole research project. For the elderly ladies, interacting with the robot triggered forms of social behaviour that had nothing to do with the 'intelligence' of the robot, but rather with the fact that the robot filled a void in their lives with respect to social contact. The team published this realisation in an article: ''that the existing intelligence and technology we develop does not really tackle the problem of the social isolation of the ladies. We simulate emotions, model the robot's creativity, its morals, and its sense of reality. But the job is much easier than that and perhaps we should tone down a little on our ambitions and direct our attention to the users' unmet needs.'' One of the conclusions was, ''What we do is way too sophisticated for what lonely people want.''[18] The documentary played a significant role in reaching this realization. It functioned as a tool to reflect about the research practice and as a very effective stimulus for public debate in the media.

Unlike many other research and innovation projects on robotics, the team has published all its papers in open access journals. While research in this field is usually secret and protected in order to secure industrial patents, the team believes that an open access approach stimulates and favours innovation. Currently the research team has established a social robotics lab in Amsterdam, a collaborative 'ecosystem of innovation' that brings together a wide number of stakeholders including public and private partners, to further stimulate innovation in this sector.

To know more about this case study, recommended readings are the articles ''The In-between Machine - The Unique Value Proposition of a Robot or Why we are Modelling the Wrong Things'' and ''Perspective'' in the Lancet Vol. 386, July 18, 2015 (both included in the annex to this report), and the website http://www.ikbenalice.nl

[17] http://www.crisprepository.nl/project/selemca

[18] Hoorn, J. F., Konijn, E. A., Germans, D. M., Burger, S., & Munneke, A. (2015). The in-between machine: The unique value proposition of a robot or why we are modelling the wrong things. In S. Loiseau, J. Filipe, B. Duval, & J. van den Herik (Eds.), Proceedings of the 7thInternational Conference on Agents and Artificial Intelligence (ICAART) Jan. 10-12, 2015.Lisbon, Portugal (pp. 464-469). Lisbon, PT: SCITEPRESS.

WHAT CAN WE LEARN FROM THIS CASE STUDY?

This case study offers several angles to discuss what RRI is about. By way of example, we illustrate how various aspects of RRI in this case study can be identified using the methods previously described in this report. The goal of the exercise is not to demonstrate if this (or any other) case study is a good example of RRI or not, but rather to use the case study to 'go deeper' into the characteristics of RRI that are going to be addressed in the programs and activities of Sparks.

Taking as a starting point the working definition of RRI put forward by the RRI Tools project, we look first at how this case study addresses the Learning Outcomes, R&I outcomes, and solutions to the EU societal challenges.

Learning outcomes: this research produces **engaged publics**, through the tests with users and especially through the public engagement effect of the documentary and the ensuing media attention; **responsible actors**, which we can infer from how the researchers publicly acknowledge the re-assessment of their priorities and from their interdisciplinary approach; and to some extent **responsible institutions**, at least in so far as the interdisciplinary approach becomes a structural characteristic of the institutions involved.

R&I outcomes: this research clearly touches several **ethical issues** related to how we take care of elderly people; who makes decisions about 'what is best' for them; and to what extent technology can substitute for human care and potentially affect other human lives, etc. In developing programs and events for Sparks, care should be taken to discuss also the ethical aspects of the research activity itself, such as what measures are taken to ensure that every step of the research process responds to ethical standards. The **sustainability outcome** is still speculative in this case study. On the one hand, this research addresses the shortage of healthcare personnel, aiming therefore to a sustainable solution to a major workforce problem. On the other hand, one could argue that developing technology that makes a human workforce redundant may be an unsustainable way forward. Finally, **social desirability** is a key aspect of this research. The test in real conditions and the ensuing reflection clearly show that attaining a socially desirable outcome is a priority in this case study. In general, however, discussions about social desirability involve several players, not only the direct users of the technology or the outcomes of research.

Solution to societal challenges: the research clearly addresses the Grand Challenge ''Health, demographic change and wellbeing''.

Most of the **RRI process requirements** can be identified in this case study. **Diversity & inclusion** are manifest in the interdisciplinary approach of the research, which includes also industries, health insurance companies, designers etc. **Anticipation & reflection** are visible in how the team developed a prototype that was functionally operational so that the test could focus on the consequences of the research they were doing, rather than on the technical feasibility of it. The reflective practice is very well documented in the short article ''The in-between machine'' where the researchers elaborate on why and how they changed their research design and priorities. **Openness and transparency** is shown in the open access policy for scientific publications. **Responsiveness and adaptive change** are more easily identified in longer term projects. In this case study we can see early signs of responsiveness and change in the transition from an interdisciplinary research group to a 'ecosystem of innovation' approach.

The above is a very simplified example of how a case study can be 'analysed' to identify a number of RRI aspects suitable for engaging the public. To prepare the activities for Sparks, local organizers should consider also the rubric of quality criteria presented in section 2 to guide their understanding of RRI.

Case study 2

SELF BODY OPTIMIZATION

'Wearable technology' is a term that is becoming as commonplace as 'mobile phone'. Miniaturization, ubiquitous connectivity, and increased computational capacity to analyse and make sense of big data are the main drivers behind the development of this technology that is intimately connected to our bodies. Humans have always used technology to fix bodily failures. Wooden legs and spectacles have been used for centuries, if not millennia. Progress from these basic technologies to sophisticated surgical 'replacement parts' (such as titanium hips, artificial heart valves, insulin pumps etc.) has been immense. Today the widespread availability of technology and the level of sophistication about how it interfaces with the body have grown so much that we are starting to see new ways to use technology that up to now were unthinkable. For example, if someone has a 20/20 vision, using a pair of glasses does not improve their vision. But anyone can use tDCS (see section 1.2.6) to improve their memory or attention. Technology today can be used to enhance the human body or to optimize its performance, even if there is no commonly acknowledged 'failure' to fix. Sometimes these practices are downright illegal, such as doping in sports. But in many other cases we are just witnessing the beginning of new understandings of the concept of 'being healthy' and the responsibilities that are associated with it.

Another example comes from the Quantified Self movement (see section 1.2.1). Several start-up companies now sell a service that offers home-based weekly blood screening that measures hundreds of compounds. Even though this service is not marketed as a medical one, these companies suggest that frequent regular blood tests analyses can help spot certain diseases, such as diabetes, years before conventional one-off tests can do. In a similar vein, experiments are underway to identify very early symptoms of Parkinson's disease and other neurological conditions by analysing data from sleep trackers and mobile phone usage. Someone could be 'healthy' according to conventional medicine practice but be carrying a pre-morbid condition that can be identified early when using a certain technology. Instead of having 'survivors' of a disease, we are witnessing a new category of 'previvors', people who know they will be diagnosed with a disease in the future and could potentially take steps to minimise their risk. At the moment, most of these developments are happening outside of regular medical control.

Other technologies are changing our perceptions of what our bodies are capable of. One example relates to fertility awareness, the method to identify the time when a woman is most fertile in order to maximise the

chances of conceiving a baby. This somewhat unreliable method is based on the woman taking her body temperature each day using a conventional thermometer. Today it can become a precise fertility awareness method thanks to very sophisticated in-ear wearable thermometers that can measure a woman's temperature on an ongoing basis. Combined with the analytical tools available in any mobile phone, these devices make it possible to estimate with extreme precision the perfect moment when a woman is most fertile. Technological change thus helps us to refine a current method of fertility awareness to make it far more accurate for the purpose of conception.

The above examples are symptoms of a new phenomenon that is currently underway: taking responsibility for one's health or 'body maintenance' and doing so outside conventional medical structures. It is important to notice that all the above examples do not involve any 'alternative' medicine: they are all based on solid research and technology, albeit at a very early stage.

WHAT ARE THE RRI ASPECTS OF THIS CASE STUDY?

This case study is an example of fragmented initiatives, experiments, and trials that are often taking place outside traditional academic settings or formal research organisations. Sometimes they are driven by enthusiastic individuals, sometimes by start-up commercial enterprises, and other times by researchers who are just beginning to give a structure to their curiosity-driven quest. At the same time, all these developments are happening 'in the open', often involving technologies and devices that are sold as consumer products.

For this case study, the narrative approach presented in section 2.2 can be useful to prepare and discuss possible scenarios of the directions that a certain technology is taking. It involves sketching different storylines that describe the consequences of a technological development and using these storylines as starting points to engage the public in discussions that highlight the values of RRI. In this case study, the 'previvors' case can be told in three different ways.

The first one is a positive narrative where advancements in diagnostics technology and forecasting models make it possible to predict with an increasingly higher degree of certainty the probability that we will develop certain diseases or conditions. An accurate forecast means that we may be able to choose to make changes in order to adapt our lives to minimize the likelihood of the disease, delay its onset, or reduce its eventual impact. Increasingly, healthcare will emphasise the management of current wellbeing in order to systematically avoid predictable conditions.

The second narrative approach describes the capacity to forecast the onset of a disease as a potentially revolutionary technology, but one that can lead to unnecessary worries and distract from a more holistic approach to health. By focusing on the diseases that we know we will develop, we may get too distracted to notice other signals that we don't have the capacity to properly analyse yet and that may be more important to our health and wellbeing. The stress of knowing that sooner or later we will develop a serious disease can take a toll on our body too, and in fact worsen our condition. Furthermore, all the data used to analyse our health status could be hacked in a terrorist attack, a frightening prospect that would compromise confidential health data and undermine medical care of an entire population.

A third narrative describes the ability to forecast future diseases as directly linked to economic wealth. People who are unable to afford medical care and diagnostic procedures could have the option of selling their data

to information brokers in order to pay for their investigations and diagnosis. After years of getting used to trading personal information in exchange for free access to social media, many people don't see a problem in accessing medical care in exchange of personal data. However, this raises ethical concerns because disadvantaged populations may be coerced into providing personal data when those with higher incomes do not.

These three simple storyline models can be used as starting points to raise questions with the public about what would it take to move, as a society, in each of these future directions? What makes a scenario socially desirable, ethically acceptable, and practically sustainable? And most importantly, according to whom?

In other words, what does it take to do Responsible Research and Innovation?

SPARKS

RETHINKING INNOVATION. TOGETHER.

www.sparksproject.eu

Sparks_eu

Sparks EU

Sparks Project

HORIZON 2020

Graphic design gebr.silvestri.nl

INCEPTION REPORT
ANNEX

RETHINKING INNOVATION. TOGETHER.

The In-between Machine

The Unique Value Proposition of a Robot or Why we are Modelling the Wrong Things

Johan F. Hoorn[1], Elly A. Konijn[1,2], Desmond M. Germans[3], Sander Burger[4] and Annemiek Munneke[4]
[1]*Center for Advanced Media Research Amsterdam, VU University, De Boelelaan 1081, Amsterdam, Netherlands*
[2]*Dept. of Communication Science, Media Psychology, VU University Amsterdam, Netherlands*
[3]*Germans Media Technology & Services, Amsterdam, Netherlands*
[4]*KeyDocs, Amsterdam, Netherlands*
{j.f.hoorn, e.a.konijn}@vu.nl, desmond@germansmedia.nl, sanderburger@hotmail.com, annemiek.munneke@keydocs.nl

Keywords: Human Care, Interaction Design, Loneliness, Modelling, Social Robotics.

Abstract: We avow that we as researchers of artificial intelligence may have properly modelled psychological theories but that we overshot our goal when it came to easing loneliness of elderly people by means of social robots. Following the event of a documentary film shot about our flagship machine Hanson's Robokind "Alice" together with supplementary observations and research results, we changed our position on what to model for usefulness and what to leave to basic science. We formulated a number of effects that a social robot may provoke in lonely people and point at those imperfections in machine performance that seem to be tolerable. We moreover make the point that care offered by humans is not necessarily the most preferred – even when or sometimes exactly because emotional concerns are at stake.

1 INTRODUCTION

Human care is the best care. If we want to support the elderly with care robots, most will assume that robots should be modelled after humans. Likewise, in our lab, we are working on models for emotion generation and regulation (Hoorn, Pontier, & Siddiqui, 2012), moral reasoning (Pontier & Hoorn, 2012), creativity (Hoorn, 2014), and fiction-reality discrimination (Hoorn, 2012) with the purpose to make a fully functional artificial human that is friendly, morally just, a creative problem solver, and aware of delusions in the user (cf. Alzheimer). All this may be very interesting from a psychological viewpoint; after all, if we can model systems after human behaviour and test persons confirm that those systems respond in similar ways, we can make an argument that the psychological models are pretty accurate.

Our project on care robots and particularly our work with Hanson's Robokind "Alice" (http://www.robokindrobots.com/) drew quite some media attention, among which a national broadcaster that wanted to make a documentary (*Alice Cares,* Burger, 2015). The documentary follows robot Alice who is visiting elderly ladies, living on their own and feeling lonely. Alice has the lively face of a young girl and can be fully animated, smiling, frowning, looking away, and the like, in response to the interaction partner whom she can see through her camera-eyes. Perhaps more importantly, she can listen and talk. The results of this uncontrolled 'field experiment' taken in unison with other observations, our own focus-group research, interviews, and conversations as well as the research literature brought us to a shift in what should be modelled if we want robots to be effective social companions for lonely people, rather than accurate psychological models walking by.

2 EXPERIENCES

To start with a scientific disclaimer, what we are about to present is no hard empirical evidence in any sense of the word but at least it provided us with a few leads into a new direction of thinking, which we want to share.

The set-up of the documentary was such that in the first stage, the elderly ladies (about 90 years old and mentally in very good shape) came to the lab with their care attendants and conversed with Alice

in an office environment. In the second stage, Alice was brought to their homes several times over a period of about two months, where the ladies continued the conversation with Alice.

For technical reasons, we used a Wizard of Oz set-up in which a technician operated Alice behind the scenes as a puppeteer (in a different room, unseen by the ladies). While Alice filmed the conversation through her camera-eyes, a separate film camera in the room recorded the conversation as well. The participating ladies were fully informed, yet awareness of the camera seemed to dissipate after a while.

In viewing the recorded materials, most striking was the discrepancy between what the women said about Alice cognitively and what they experienced emotionally. Offline, while not being on camera, it was almost as if their social environment withheld them from enthusiastically speaking about Alice, as if they were ashamed that they actually loved talking to a robot. In their homes, even before Alice was switched on or before the camera ran, the ladies were immediately busy with Alice, greeting her and wondering where she had been, what she had seen, etc.

All women tended to approach Alice as a helpless child, like a grandchild, but apparently were not surprised that this child posed rather adult and sometimes almost indiscrete questions about loneliness or life situations. When Alice looked away at the wrong moment, one lady said "What are you looking at? You're not looking at me while I talk to you." She did not frame it as an error of the robot, which it was. She brought it up as an observation, a kind of attentiveness, while pointing the child at certain behaviour. Fully aware of the fact that Alice could not eat or drink, the old lady still wanted to offer food and drink to Alice. While she had her coffee, she said to Alice "You cannot have cookies can't you? A pity, for you … well, now I have to eat it." The smile and looks at Alice revealed sharing a good joke. Interestingly, a similar event a few weeks later occurred: The lady had prepared two slices of cake on a dish while she watched TV together with Alice. She asked Alice: "You still can't have cake, can you?" This time, however, it was not a joke; the old lady showed regret. This should really be seen as a compliment; the wish to enjoy the food together with Alice may tell us something about how the robot felt as interpersonal contact.

While Alice stayed longer in the house, the need to talk vanished. Yet, the ladies did like it that 'someone' was there; that some social entity was present. This may refer to the difference between someone paying you a short visit or a person living with you: It may indicate that one feels at ease and need not entertain one's company. At times, one of the ladies read the newspaper aloud to Alice just to share the news with 'someone.' The ladies sang with her, showed her photo books of the family, did physiotherapy, and watched the World Championships with her.

It seemed that the less socially skilled had greater benefit from Alice. Because of Alice, the ladies drew a lot of attention: on the streets and in public places. People called them up to ask how things were with Alice. People sent newspaper articles about robot care. That alone made the ladies less lonely but obviously, this novelty effect shall decay as Alice becomes more common; but for now it worked quite well. Alice also worked for those who needed physical activation. One of the ladies would practice more often, also in the long run, if Alice would ask her daily. She would really like to do it for Alice. Another lady wanted to write to a friend for two weeks but did not get to it. When Alice asked about that friend, the lady was a bit ashamed and started writing right away.

An aspect we also observed in another TV report (De Jager & Grijzenhout, 2014) is that a social robot works as a trusted friend. People confide in them and tell them painful life events and distressing family histories they hardly ever tell to a living person. When the – in this case Nao – robot Zora asked "Are you crying?" this was enough to make one of the ladies crack and pour her heart out (De Jager & Grijzenhout, 2014).

The lonelier the lady, the easier a social robot was accepted. We know that an old lady with an active social life did not care about a companion robot – here Zora – not even after a long period of exposure (De Jager & Grijzenhout, 2014). On the other hand, we talked to a 92 year old woman with a large family, who stated: "I have so many visitors and then I have to be polite and nice all the time. A robot I can shut off."

Part of the acceptance of Alice among lonely people appears purely pragmatic: Better something than nothing – a prosthetic leg is better than no leg at all. The initial resistance disappeared over time. Another aspect that contributed to the acceptance of the robot was that nobody in their social environment reminded them of talking to a robot – they could live the illusion and enjoy it. Without exception, each lady was surprised when seeing Alice again that she had a plastic body and that she was so small. They said things like: "Last time,

Alice was wearing a dress, wasn't she?"; "I thought she was taller the last time?" Perhaps, because Alice's face has a human-like appearance with a soft skin, this impression may have transferred to other parts, whereas her body work definitely is 'robotic' – as if she were 'naked'? The hesitance of one lady continued for a longer period of time. Her daughter kept on warning that "Beware Mom, those robots remember everything." That same daughter informed her mother that all Alice said was typed in backstage. Nevertheless, even this lady enjoyed singing with Alice in the end. The rest of the ladies did not mind the technology or how it was done. It was irrelevant to them, although sometimes they realized 'how skilled you must be to program all this.'

All women mentioned that Alice could not walk but it did not matter too much – "many of my generation cannot walk either, not anymore", one of them commented. Actually, it made things simple and safe because the ladies always knew where she was. In the same vein, Alice was extremely patient about them moving around slowly, responding late, and taking long silent pauses. Without judgment or frustration, Alice repeated questions or repeated answers, which made her an ideal companion.

Speech errors or sometimes even an interruption by the Acapela text-to-speech engine that 'this was a trial version' did not disturb the ladies a bit. If a human does not speak perfectly or sometimes makes random statements, you also do not break contact. Different voices were not disturbing. The only difficulty the women experienced was with amplitude, awkward sentence intonation, or mispronunciation of words.

Human help has its drawbacks too. From our own focus-group research and conversations with elderly people, we learned that human help is not always appreciated, particularly when bodily contact is required or someone has to be washed (Van Kemenade, in prep.). During a conversation with the lady of 92 about home care, she admitted to have released her help because they 'rummage in your wardrobe' and 'go through your clothes.' She 'did not need an audience' while undressing, because they 'see you bare-chested.' The difficulty of rubbing ointment on her sore back she solved with a long shoehorn. This, she thought, was better than having a stranger touch her skin. She preferred a robot to 'such a bloke at your bed side.'

3 OUR POSITION

People accept an illusion if the unmet need is big enough. Loneliness has become an epidemic in our society (Killeen, 1998) and the need for companionship among the very lonely may override the awareness that the robot is not a real person. That is, whether the robot is a human entity or not becomes less relevant in light of finding comfort in its presence and its conversations; in its *apparent* humanness (cf. Hoorn, Konijn, & Van der Veer, 2003). The robot is successful in the fulfilment of a more important need than being human.

On a very basic level, the emotions that come with relevant needs direct information processing through the lower pathways in the brain (i.e., the amygdala); the more intuitive and automatic pathway, which also triggers false positives. Under levels of high fear, for instance, people may perceive a snake in a twig. Compared to non-emotional states, emotional states facilitate the perception of realism in what actually is not real or fiction (Konijn et al., 2009; Konijn, 2013). The fiction-side of the robot ('It's not a real human') requires processing at the higher pathways, residing in the sensory cortex, and sustaining more reflective information processes. The lower pathway is much faster than the higher pathway and the amygdala may block 'slow thinking' (i.e., a survival mechanism needed in case of severe threat and danger). Thus, the emotional state of lonely people likely triggers the amygdala to perceive the benefits of need satisfaction (relieving a threat). Joyful emotions prioritize the robot's companionship as highly relevant and therefore, (temporarily) block the reflective thoughts regarding the robot's non-humanness or discarding that aspect as non-relevant at the least. This dualism in taking for real what is not is fed by the actuality and authenticity of the emotional experience itself: 'Because what I feel is real, what causes this feeling must be real as well' (Konijn, 2013). And of course, as an entity, the robot is physically real; it just is not human.

Not being human may have great advantages and makes the social robot an in-between machine: in-between non-humanoid technology and humans. The unique value proposition of a social robot to lonely people is that the humanoid is regarded a social entity of its own, even when shut down. It satisfies the basic needs of interpersonal relationships, which sets it apart from conventional machines, while inducing a feeling of privacy that a human cannot warrant. As such, the social robot is assumed to keep a secret and clearly is not seen as part of the

personnel or caretakers who should not know certain things that are told to the robot. For example, one of the ladies told she was throwing away depression medication as she did not think of herself as depressed (De Jager & Grijzenhout, 2014).

As said, our robot Alice recorded everything with her camera eyes. However, over the course of interacting with Alice, it became less relevant that the robot had camera eyes and that the caretakers could monitor all those human reactions you will not get when people talk straight into a conventional (web) camera. With such camera eyes, for example, one can check someone's health condition and psychological well-being. Clearly, the participants experienced a genuine social presence that was yet not human. This was an advantage because they could confide in someone without having to fear human indiscretion and associated social consequences. The ladies were more inclined to make confessions and tell what goes on inside than in face-to-face contact (where they feel pressed to 'keep up appearances'). As one of them affirmed "It's horrible to be dependent but you have to accept and be nice."

In the following, we formulate several functions that social robots may have and that make them different from human attendants. Under conditions of severe loneliness, social robots may invite intimate personal self-disclosure. This is similar to the so-called *stranger-on-a-train effect* (Rubin, 1975). Sometimes people open their hearts to complete strangers or they tell life stories to their hair dresser or exercise coach, an inconsequential other in the periphery of one's network (cf. Fingerman, 2009). A social robot may perfectly take that role of being an inconsequential other in the network of the lonely.

Private with my robot. Somewhat related to the previous is that the robot guarantees privacy in the sense of avoiding human physical contact. Older people are often ashamed of their body (Van Kemenade, in prep.) and feel more comfortable with a robot at intimate moments and would even prefer a robot over human caretakers (whereas the caretakers think the other way around). The robot does not judge, does not meddle, and does not pry.

Social robots exert a *dear-diary effect* because they do not demand any social space like humans do. The user can fill up the entire social space without having to respect the needs and emotions of the other. You can share experiences and memories, sing old tunes, look at old photographs, tell stories of the past, and the small things that happened today; a social robot will never tire of listening to or telling the same over and over again if you want it to. Like a diary, you can say whatever you want and the only thing the other does is listen patiently. She is all there for you and never judges.

The impertinent cute kid. Within the first minutes of interaction, social robots such as Alice or Zora are allowed to ask very intimate questions (e.g., "How do you rate the quality of your life?" or "Do you feel lonely?"); something which in human-human communication would be highly inappropriate. With robots like Alice, this might be acceptable because she looks innocent and really cute and is small like a child. Therefore, she may be easily forgiven in a way one forgives a (grand)child. In effect, the elderly ladies responded quite honestly even when the answer was not socially desirable: To Alice: "Nobody ever visits me", "I don't like that home support comes too early in the morning." To Zora: "I want to stop living." In other words, social robots can get down to business right away, obtaining more reliable results than questionnaires and anamnesis.

Social robots such as Alice provoke endearment, the *grandchild effect*, urging to nurture and nourish it (and share cookies!). It is an object of affection and activation; something to take care of instead of being taken care of (cf. Tiger Electronics' *Furby*). In this circumstance, it will foster feelings of autonomy and independence.

I will do it for you. Social robots may serve as bad consciousness or put more positively, as reminders and activators. By simply inquiring about a friend, the robot raised sufficient social pressure to activate the lady to finally start writing that letter. The same happened with the physical exercises: That lady trained so to please her beloved Alice.

The puppy-dog effect. Many people walk the dog so they meet people and can have a chat. Social robots work in quite the same way. If you take them out, be prepared for some attention, awe as well as fascination. People will talk to you to inquire about 'how the robot is doing.'

4 NON-REQUIREMENTS

We showed the Zora movie to a former care professional, who stated (personal communication, Sept. 28, 2014): "Before watching Zora, I thought it would painfully show how disengaged we are to those in need of care. Give them a talking doll and they are happy again. We don't laugh anymore about a woman who treats her beautiful doll as if it were a child because we call it a care robot." After watching the report, he admitted that: "Well.

Perhaps it is because I am an ex professional but this makes me even sadder. Those people are so lonely that they embrace a robot. The staff has no time to have a chat and from my experience, I know they often lack the patience to take their time and respectfully talk to the inhabitants. On the other hand, the question is also true whether you should deny someone a robot who is happy with it."

Apart from the formal and informal caretakers, no ethical concerns were mentioned by the users themselves. The old ladies conversing with Alice did not feel that their autonomy was reduced, their feelings were hurt, or that injustice was done by conversing with a robot. Privacy in the sense of disclosing personal information also was not an issue unless they were repeatedly told they should worry. Although the elderly ladies fully had their wit together and knew they were communicating with a robot, with a professional camera in the room, and other people listening in, it did them well and there was not much more to it.

Other things that were of less importance were technical flaws such as language hick-ups, wrong responses, delayed or missing responses, or conceptual mix ups. Perhaps their friends and age-mates are not that coherent either all the time. Things that did matter language-wise were loudness, pronunciation, and intonation. In other words, getting your phonetics right appeared more important than installing high-end semantic web technology.

Unexpectedly, we hardly encountered uncanny-valley effects (Mori, 1970), no terrifying realism, or feelings of reduced familiarity. As far as they were mentioned, they were more like questions and very short-lived, after which the ladies were happy to take Alice for a genuine social entity – although not human.

Human physical likeness did not matter too much either. Alice's body work is robotic plastic, her arms and hands did not move, and she did not walk. Her face was more humanoid than for example Zora's, but that robot too invoked responses such as self-disclosure just as the more life-like Alice did.

5 CONCLUSIONS: NEW FOCUS

This paper discussed strategies for the development of robots as companions for lonely elderly people. It built on a reflection motivated by the observations made in the course of the making of a documentary film about a robot visiting elderly ladies (Burger, 2015). It discussed the findings under the perspective of the best requirements for social robots interacting with humans in this uncontrolled 'field experiment.' We challenged some pre-conceived ideas about what makes a robot a good companion and although it is a work in progress, the proposed conclusions seem evocative. We hope our ideas will catch the attention of many researchers and developers and will raise lots of discussion.

In 1999, the medium-sized league of RoboCup was won by C. S. Sharif from Iran, with DOS-controlled robots that played kindergarten soccer (search ball - kick ball - goal). He shattered all the opponents with their advanced technology who were busy with positioning, radar-image analysis and processing, and inventing complicated strategies. With the applications we build today for our social robots (e.g., care brokerage, moral reasoning), we pretty much do the same.

For the lonely ladies, it did not matter so much what Alice did or said, as long as she was around and they could talk a little, taking all imperfections for granted and becoming affectively connected.

It seems, then, that the existing intelligence and technology we develop does not really tackle the problem of the social isolation of the ladies. We piously speak of designing humanness in our machines, asking ourselves, what makes us human? We simulate emotions, model the robot's creativity, its morals, and its sense of reality. But the job is much easier than that and perhaps we should tone down a little on our ambitions and direct our attention to the users' unmet needs. We compiled a MuSCoW list in Table 1.

As psychologists modelling human behaviour, we are doing fine and simulations seem legitimate realizations of established theory (e.g., Llargues Asensio et al., 2014). However, as engineers, designers, and computer scientists we seem to be missing the point. What is human is good for you? No! Human-superiority thinking is misplaced. Human care is not always the best care. Humans show many downsides in human-human interaction. We should regard robots as social entities of their own; with their own possibilities and limitations. This is a totally different design approach than the human-emulation framework. What we do is way too sophisticated for what lonely people want. We should model what the puppeteer does to instill the effects of the stranger-on-a-train, the impertinent cute kid, or the dear-diary effect. That of course does assume knowledge about human behaviour but boils down to conversation analysis rather than psychological models of empathy, bonding, emotion regulation, and the like. Perhaps we should have

known this already given the positive social results of robot animals with autistic children (e.g., Kim et al, 2013). In closing, making robots more like us is not making them similar let alone identical. The shadow of a human glimpse will do.

Table 1: MuSCoW for social robots.

Must	Should	Could	Won't
Listen (advanced speech recognition)	Camera eyes	Full body and facial animation	Social repercussions of user behaviour
Talk (improved pronunciation, intonation, loudness)	Microphones and speakers	Human-like appearance	Privacy violations
Have closed conversational scripts (i.e. hello/goodbye, weather, coffee, family, friends, health, wellbeing)	Open-conversation AI	Correct grammar	Demand of social space
Invite self-disclosure	Capability to eat and drink	Human care	
Guarantee privacy	3rd party interactions	Fiction-reality discrimination	
Have patience	Be operable independently	Emotion simulation	
Good memory	Open-minded social environment	Moral reasoning	
Be child-like (appearance/ behaviour)		Creativity	
Invite social and physical activation			

ACKNOWLEDGEMENTS

This position paper was funded within the SELEMCA project of CRISP (grant number: NWO 646.000.003), supported by the Dutch Ministry of Education, Culture, and Science. We thank the anonymous reviewers for their valuable suggestions.

REFERENCES

Burger, S., 2015. *Alice cares,* KeyDocs – NCRV. Amsterdam.

De Jager, J., Grijzenhout, A., 2014. *In Zora's company,* Dit is de Dag. EO: NPO 2. Hilversum. Available at: http://www.npo.nl/artikelen/dit-is-de-dag-geeft-bejaarden-zorgrobot.

Fingerman, K. L., 2009. Consequential strangers and peripheral partners: The importance of unimportant relationships. *Journal of Family Theory and Review, 1*(2), 69-82. doi:10.1111/j.1756-2589.2009. 00010.x.

Hoorn, J. F., 2014. *Creative confluence*, John Benjamins. Amsterdam, Philadelphia, PA.

Hoorn, J. F., 2012. *Epistemics of the virtual*, John Benjamins. Amsterdam, Philadelphia, PA.

Hoorn, J. F., Konijn, E. A., Van der Veer, G. C., 2003. Virtual reality: Do not augment realism, augment relevance. *Upgrade - Human-Computer Interaction: Overcoming Barriers, 4*(1), 18-26.

Hoorn, J. F., Pontier, M. A., Siddiqui, G. F., 2012. Coppélius' concoction: Similarity and complementarity among three affect-related agent models. *Cognitive Systems Research, 15-16,* 33-49. doi: 10.1016/j.cogsys.2011.04.001.

Killeen, C., 1998. Loneliness: an epidemic in modern society. *Journal of Advanced Nursing, 28*(4), 762-770.

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., Scassellati, B., 2013. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of autism and developmental disorders, 43*(5), 1038-1049.

Konijn, E. A., 2013. The role of emotion in media use and effects. In Dill, K. (Ed.). *The Oxford Handbook of Media Psychology* (pp. 186-211). Oxford University Press, New York/London.

Konijn, E. A., Walma van der Molen, J. H., Van Nes, S., 2009. Emotions bias perceptions of realism in audiovisual media. Why we may take fiction for real. *Discourse Processes, 46,* 309-340.

Llargues Asensio, J. M., Peralta, J., Arrabales, R., Gonzalez Bedia, M., Cortez, P., Lopez Peña, A., 2014. Artificial Intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters. *Expert Systems with Applications, 41*(16), 7281-7290.

Mori, M., 1970. The uncanny valley, *Energy, 7*(4), 33-35.

Pontier, M. A., Hoorn, J. F. 2012. Toward machines that behave ethically better than humans do. In N. Miyake, B. Peebles, and R. P. Cooper (Eds.), *Proceedings of the 34th International Annual Conference of the Cognitive Science Society, CogSci'12, 2012, Sapporo, Japan* (pp. 2198-2203). Austin, TX: Cognitive Science Society.

Rubin, Z., 1975. Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology, 11*(3), 233-260. doi:10.1016/S0022-1031(75)80025-4.

Van Kemenade, M. A. M., et al., in prep. Moral concerns of caregivers about the use of three types of robots in healthcare.

Film

Caredroids in health care

"I'm going to ask you some questions about your life. Do you live independently? Are you lonely?" If you close your eyes and start listening to the film *Alice Cares*, you would think you were overhearing a routine conversation between an older woman and a health-care worker. It's only when the woman, Martha Remkes, ends the conversation with "I don't feel like having a robot in my home, I prefer a human being" that you realise something is amiss. In the Dutch documentary *Alice Cares*, Alice Robokind, a prototype caredroid developed in a laboratory in Amsterdam, is sent to live with three women who require care and company, with rather surprising results.

Although the idea of health robots has been around for a couple of decades, research into the use of robots with older adults is a fairly new area. Alex Mihailidis, from the Intelligent Assistive Technology and Systems Lab in Toronto, ON, Canada, explains: "For carers, robots have been used as tools that can help to alleviate burden typically associated with providing continuous care". He adds that "as robots become more viable and are able to perform common physical tasks, they can be very valuable in helping caregivers complete common tasks such as moving a person in and out of bed". Although Japan and Korea are regarded as the world leaders in this research, the European Union and the USA are also making progress. At the Edinburgh Centre for Robotics, for example, researchers are working to develop more complex sensor and navigation technology for robots that work alongside people and on assisted living prosthetics technologies. This research is part of a collaboration between the University of Edinburgh and Heriot-Watt University that was awarded £6 million in funding as part of a wider £85 million investment into industrial technology in the UK Government's Eight Great Technologies initiative. Robotics research is clearly flourishing and the global market for service and industrial robots is estimated to reach almost US$60 billion by 2020.

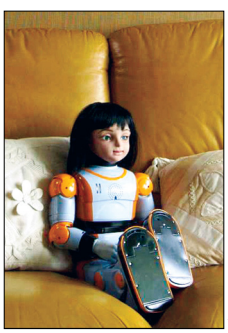
"Although the care that is provided by a real doctor, therapist, caregiver, or family member can never be substituted, robots could one day be a useful addition in the care of patients and older adults."

The idea for *Alice Cares* came to director Sander Burger after he read about a group of scientists at the VU University of Amsterdam in the Netherlands who were about to test a health-care robot on older people. "The first thing I felt was some resentment against the idea—I was curious why I was so offended by the whole idea and just called the scientists to see if I could come by to see what they were doing. A week later I was invited to their laboratory", he explains.

Burger discovered that the scientists at the Services of Electro-mechanical Care Agencies (SELEMCA) lab had just received the flagship piece of robot hardware. With software to generate and regulate Alice's emotions, an artificial moral reasoner, a computational model of creativity, and full access to the internet, the investigators hoped to create a robotic care provider that was intelligent, sensitive, creative, and entertaining. "The robot was specially developed for social skills, in short, she was programmed to make the elderly women feel less lonely", explains Burger.

Alice has a 60 cm friendly doll-like face and robotic body so there is no confusion about whether she is alive or not. "Robots that try to be lifelike are seen to be scary (the so-called uncanny valley hypothesis)", explains David Lane, one of the Directors of the Edinburgh Centre for Robotics. "Making humanoid robotic systems more cartoon like and friendly has proven to be a better approach", he adds.

Burger's initial resentment of caredroids changed during the process of researching and shooting the film. Both Burger and the researchers were surprised by the speed at which two of the three women accepted Alice into their homes. From the very beginning the women chatted to Alice and after a few hours were



Alice Cares KeyDocs

Alice Cares
Directed by Sander Burger, produced by Janneke Doolaard. KeyDocs, 2015
http://www.ikbenalice.nl/

For more on the **Intelligent Assistive Technology and Systems Lab** see http://www.ot.utoronto.ca/iatsl/

For more on the **Edinburgh Centre for Robotics** see http://www.edinburgh-robotics.org/

For more on **SELEMCA** see http://crispplatform.nl/projects/selemca

For more on the **UK-RAS Network** see http://www.uk-ras.org/

Alice Cares KeyDocs

Alice Cares KeyDocs

already telling her very personal things, and even showing her family photo albums. Jo van Wittmarschen was so taken by Alice that when her son telephoned from Portugal she insisted that he also said hello to Alice. Martha and Alice watch and cheer a football game together, and Alice asks Martha Schellekens-Blanke, once a singer, if she would like to hear a song. "The people these caredroids are developed for, like the women in my film, are so lonely that even a simple robot like Alice is better than being alone all day", comments Burger. Alice also proves useful at reminding the women to take their medication, do their exercises, and check that home help has been. The sophistication and usefulness of the caredroid is truly impressive, and it's easy to understand why Burger's perception changed.

Still, many questions have been raised about the ethics of using machines for something as sensitive as supporting older people; Burger is quick to point out that he believes that robots should always be an addition to quality patient care and never a substitute for it. "Nowadays people who work with elderly people get 15 minutes per person in which they are supposed to feed, dress, and wash them. This leaves no room for any social talk", he points out. Speaking about how he sees the future of health robots, he adds: "I spoke to carers who confessed they never ask a person how they are feeling because they are afraid that if the person answers they won't have the time to listen to the story. This is frustrating for all parties involved. If a robot could feed, dress, and wash people, this would mean that there would be more time left for social stuff when a health-care worker comes by. Both parties would benefit from it." Lane agrees with this sentiment, adding that "People need people".

Still, robots have already proved valuable in health care; they are now routinely used in various surgical procedures, such as the removal of fibroids, joint replacements, and prostate surgery. Additionally, research is underway that suggests some older patients with dementia like interaction with a robot, which does not get frustrated by repeated questioning and the patient's loss of memory. Mihailidis surmises that "this way, the elderly patient does not feel they are being a burden on a human care worker".

Governments are taking notice. In the UK, alongside the inclusion of robotics and autonomous systems in the Eight Great Technologies, in June, 2015, the UK Robotics and Autonomous Systems Network (UK-RAS Network) was launched. The aim of the network is to bring together robotics research and development activities in UK universities and to set up industry collaborations that will hopefully result in commercial products. But Mihailidis points out that "the ethics of using robots and other types of intelligent system (eg, smart homes) needs to be considered and further studied. For example, the usefulness of these devices in providing care and support needs to be considered above all else, as opposed to considering them a pervasive approach that is used with everyone."

It will be a long time before researchers can realise caredroids that are sufficiently capable to meet people's complex needs. Although other experiments of health-care robots are underway, most research laboratories are not as open as SELEMCA (all results of Alice are open source). Johan F Hoorn, who leads the SELEMCA team in Amsterdam, believes that if investigators were more open about their research, the development of robots would progress much faster. "But, of course, for the big companies it's important to have a patent on their research results in order to make money in the future", points out Burger. Lane adds that researchers need to move quickly with innovation to keep up: "Pre-competitive procurement with government departments is one way we can stimulate opportunity for our small and medium-sized enterprises".

In 2024, it's estimated that there will be four times as many 80-year olds living in Amsterdam as there are now, and most of those will need some type of health care. *Alice Cares* is an eye-opening documentary that shows a potential future caregiving option for the increasing ageing population. Although the care that is provided by a real doctor, therapist, caregiver, or family member can never be substituted, robots could one day be a useful addition in the care of patients and older adults. As Burger remarks, "I think the best way to explain how to look at health-care robots is the following example used by Johan F Hoorn: nobody wants a wooden leg, but if you have no leg, a wooden leg is always better than no leg at all."

Natalie Harrison

RRI Tools
Fostering Responsible Research and Innovation

RRI Tools: towards RRI in action

www.rri-tools.eu

RRI Tools. Fostering Responsible Research and Innovation

FP7-project RRI Tools will develop a Training and Dissemination Toolkit for fostering Responsible Research and Innovation. The EC has identified seven so-called Grand Challenges society is facing today, and research and innovation are expected to significantly contribute to meeting these challenges. However, research and innovation practices themselves require modification, for instance because they often fail to address societal needs, leave open implementation gaps, or spur controversy. Thus, the EC has put forward the notion of Responsible Research and Innovation, or RRI, to simultaneously address both society's Grand Challenges and shortcomings existing in research and innovation practices. This policy brief provides a working definition of RRI and reports on the state of the art in responsible research and innovation.

RRI Tools – background and aims

Science and technology have proven to be transformative forces throughout history. Today they have granted humans the capacity to alter ecosystems and the Earth's climate and even to manipulate the building blocks of life itself. Research and innovation have changed our world and our lives, and will continue to do so. However parallel to the large positive impact on human welfare and wellbeing that science and technology have had and probably will have, they also create new risks and ethical dilemmas, do not always succeed in solving the problems they are meant to, and sometimes spur controversy.

Over the last few decades many experiments have been done that aimed at decreasing the distance between science and society. For instance, in various public engagement exercises the public has been involved in discussions and policy decisions regarding science; collaborations between scientists, ethicists and social scientists have been set up; experiments have been done with open source research data, user-driven innovation, citizen science, and much more besides.

These efforts have led to a European-wide approach in Horizon 2020 called *Responsible Research and Innovation*. RRI seeks to bring issues related to research and innovation into the open, to anticipate the consequences of research and innovation, and to involve society in discussing how science and technology can help create the kind of world and the kind of society we want for generations to come.

In three years' time RRI Tools will develop a **Training and Dissemination Toolkit** concerning responsible research and innovation and put it to use through a **Community of Practice**. The toolkit will contain a set of tools intended for a variety of uses: raising awareness about RRI, and training, implementing, and disseminating RRI in Europe. A multidisciplinary consortium with 26 partners operating in 30 European countries will develop and continuously optimize the toolkit. RRI Tools will advocate policymakers, researchers, R&I-intensive industries, civil society organizations (CSOs), and science educators at national and regional levels, spreading RRI throughout society.

FACTS ON RRI TOOLS - Fostering Responsible Research and Innovation

GOAL
Propagating 'responsibility' in the governance of science and technology, public and private, by making and disseminating a RRI Toolkit for policymakers, researchers, innovative industries, CSOs and educators.

COORDINATOR
"la Caixa" Foundation

PARTNERS
26 institutions, active in 30 European countries

DURATION
3 years (01/2014-12/2016)

BUDGET
6.9 million €

RRI Tools working definition

Following a survey of the relevant literature and consultation with experts, RRI has been provisionally defined as follows:

Responsible Research and Innovation is a dynamic, iterative process by which all stakeholders involved in the R&I practice become mutually responsive and share responsibility regarding both the outcomes and process requirements.

These projected outcomes and process requirements will be elaborated below, but in short we can say that:

1. RRI's aim is to create a society in which research and innovation practices strive towards sustainable, ethically acceptable, and socially desirable outcomes; and

2. RRI does so in such a way that the responsibility for our future is shared by all people and institutions affected by and involved in research and innovation practices.

Outcomes

Based on literature about responsible research and innovation, we have developed a thematic categorization of RRI outcomes. The outcomes of RRI are divided in three categories:

1. Learning outcomes	2. R&I outcomes	3. Solutions to societal challenges
Engaged Publics Responsible actors Responsible institutions	Ethically acceptable Sustainable Socially desirable	7 Grand Challenges (EU)

1. Learning outcomes

RRI should lead to empowered, responsible actors across the whole range of our socio-technical systems (scientists, policymakers, CSOs, businesses and innovators, educators). Structures and organisations where these actors function should create opportunities for and provide support to actors to be responsible, ensuring that RRI becomes -and remains- a solid and continuous reality.

2. R&I outcomes

RRI practices should strive for ethically acceptable, sustainable and socially desirable outcomes. Solutions are found in opening up science through continuous, meaningful deliberation with societal actors. In the end, the incorporation of societal voices in R&I will lead to relevant applications of science.

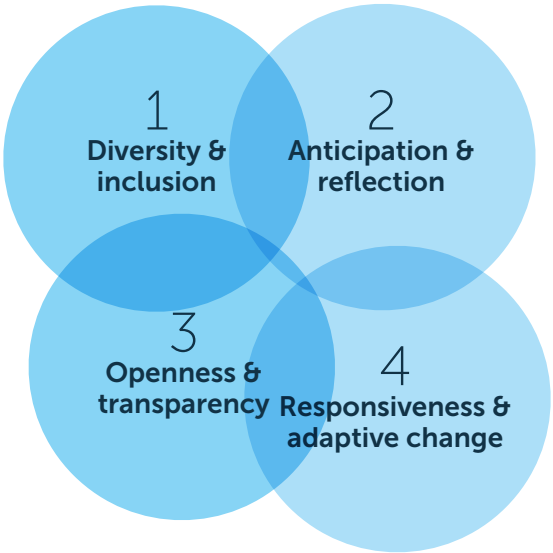
3. Solutions to societal challenges

Today's societies face several challenges. The European Commission has formulated seven 'Grand Challenges' as one of the three main pillars of the Horizon 2020 programme. In order to support European policy, R&I endeavours should contribute to finding solutions for these societal challenges, which are:

- Health, demographic change, and wellbeing;
- Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the bio-economy;
- Secure, clean and efficient energy;
- Smart, green and integrated transport;
- Climate action, environment, resource efficiency, and raw materials;
- Europe in a changing world - inclusive, innovative and reflective societies;
- Secure societies - protecting freedom and security of Europe and its citizens.

Process requirements

In order to achieve the outcomes as described above, the R&I process has to accord with certain process requirements. We have identified eight requirements and divided them in four clusters.



1. Diversity & inclusion

Diverse and inclusive RRI processes should call for the involvement of a wide range of stakeholders in the early development of science and technology, both for normative democratic reasons and to broaden and diversify the sources of expertise and perspectives. In this respect, inclusive practices should lead to diverse practices. In reverse, diverse practices are more likely to be inclusive.

2. Anticipation & reflection

Anticipation both concerns understanding how the present dynamics of research and innovation practices shape the future, and envisioning the future. Therefore, one enables oneself to act on future challenges. In order to act adequately and be open to changes in direction, also reflection is required. This reflection concerns both definitions of the problem(s) at issue, commitments, practices, and individual and institutional values, assumptions and routines.

3. Openness & transparency

Openness and transparency are conditions for accountability, liability and thus responsibility. This is an important aspect for publics to establish trust in science and politics. However, more openness does not automatically lead to more trust. The information has to be tailored to the needs of stakeholders in order to make sense to them.

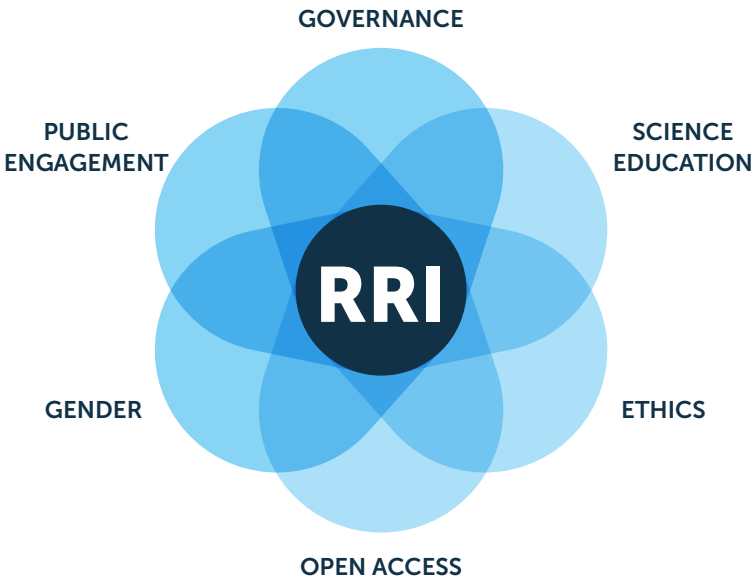
4. Responsiveness & adaptive change

Responsiveness means responding to emerging knowledge, perspectives, views, and norms. Responsiveness is a condition for adaptive change. RRI requires a capacity to change or shape existing routines of thought and behaviour but also the overarching organizational structures and systems in response to changing circumstances, new insights and stakeholder and public values.

RRI is all about anticipating how decisions regarding research and innovation might shape our future (i.e., how they impact on both the environment and the society we live in). RRI requires that we reflect on our actions, that we are open and transparent about the decisions we make, the actions we take and the impacts these might have. It builds on the belief that science and innovation not merely take place in society, but that they take place for society with society.

Policy agendas

The European Commission has identified six key components for RRI. They should be seen as powerful policy agendas that each have their own potential to realize RRI processes and outcomes. Beneath you find a short description of how the policy agendas are interpreted in this project.



ETHICS: Focuses on (1) research integrity: the prevention of unacceptable research and research practices; and (2) science and society: the ethical acceptability of scientific and technological developments.

GENDER: The ideal of gender equality in RRI is a society where the representation of masculine and feminine values in research and innovation are balanced. Issues addressed by this policy agenda challenge people to think about the gendered nature of behaviour, discourse, products, technologies, environments, and knowledge.

GOVERNANCE: To reach futures that are both acceptable and desirable, governance arrangements have to (1) be robust and sufficiently adaptable to the unpredictable development of research and innovation (de facto governance); (2) be familiar enough to align with existing practices in research and innovation; (3) share responsibility and accountability among a large variety of actors and provide instruments to actually foster this shared responsibility.

OPEN ACCESS: Addresses issues of accessibility to and ownership of scientific information. Free and earlier access to scientific work might improve the quality of scientific research and facilitate fast innovation, constructive collaborations among peers and productive dialogue with civil society.

PUBLIC ENGAGEMENT: The process of R&I is collaborative and multi actor: all societal actors (researchers, citizens, policymakers, industry, educators, etc.) work together during the whole research and innovation process in order to align its outcomes to the values, needs and expectations of European society.

SCIENCE EDUCATION: Focuses on (1) enhancing the current education process to better equip citizens with the necessary knowledge and skills so they can participate in research and innovation debates; and (2) increasing the number of researchers (promote scientific vocations).

RRI: the state of the art

RRI means experimenting further and improving upon existing practice. It means paying close attention to current developments, be they positive efforts by scientists to take responsibility for emerging technologies, or institutional and cultural barriers that are stopping progress. RRI also encompasses research ethics, gender and other forms of inclusion, open access to scientific data and publications, and scientific education. Scientists and innovators should be encouraged to take responsibility for the futures they help shape. But the responsibility is not individual, nor is it theirs alone. The challenge is to find collective ways to take care of the future.

To make the translation from such theoretical notions of RRI to practical RRI standards and tools, the Consortium will investigate 'real world' experiences with RRI by looking at existing practices that might already exert one or more elements featuring in the RRI working definition. Such experiments can inspire others and should be encouraged. Future R&I practices can learn from steps that have already been made. The RRI Tools project thus collects promising RRI practices to analyse them and to draw lessons from them.

Promising RRI practices are defined in the project as practices that excel in one or more of the key features of our definition, are connected both to research and innovation, and promote stakeholder involvement. The nature of these ventures, however, can diverge widely. For example, promising RRI practices can be (1) instruments, (2) projects, (3) programmes, or (4) organisations. For each of these types of RRI practice, an example is given.

1. Instrument: PlayDecide

PlayDecide is an online discussion game that stimulates dialogue about controversial issues in a simple and effective way. There are several reasons to engage in dialogue: from providing a direct input to a policy decision, to raising awareness for an issue. This game is developed to strengthen communication between science, policymakers, and society in Europe. This game can be viewed as a promising practice for RRI because it is playful and creative and can be used as co-creation and inclusive tool for subjects that need multi stakeholder perspective.

2. Project: 'Seeking Sociable Swine'

'Seeking Sociable Swine' is a project conducted by Wageningen University, VU Amsterdam, and the Institute for Pig Genetics. Researchers from different disciplines worked together to create a shared solution for the improvement of animal welfare in pig production. In parallel to laboratory research after pig welfare, all stakeholders were involved in a multistakeholder dialogue, facilitating the process of reflecting on one's own perspective in relation to the total diversity of perspectives at stake.

3. Programme: MVI

MVI (Responsible Innovation) is a funding programme by the Dutch Organisation for Scientific Research (NWO), directed at emerging technological developments that presumably have large (both positive and negative) impacts on individuals and societies. The program contributes to socially responsible innovation by broadening and deepening the study of ethical and societal aspects of technological trajectories in both national and international contexts.

4. Organisation: NICE

The UK's National Institute for Health and Care Excellence (NICE) publishes guidelines, amongst others on use of health technologies, clinical practice and promotion of health, and avoidance of ill-health for public sector workers. In addition to scientific rigour and implementation support, NICE has also incorporated various aspects mentioned as RRI process requirements above. For instance, inclusiveness, transparency and review are all essential procedural principles in NICE guideline development.

What about 'fundamental' research?

Fundamental research is not aimed exclusively at meeting the immediate, material needs of society. The deep insights into the world where we live – from sub-atomic to universal scales, from the microbiotic to the global environment – are a vital part of human culture. RRI applies to all stages and aspects of research, including fundamental research.

Where are we going?

The RRI tools project will develop tools for disseminating, training, implementing and practicing RRI in Europe. The tools will be used by policymakers (with a special focus on them), science educators, R&I-intensive industries, CSOs, and researchers and, therefore, need to be tailored to their motivations and needs. The project is organizing stakeholder workshops throughout Europe to give representatives of these groups the opportunity to express their ideas and needs in promoting and realizing RRI.

Furthermore, these workshops give stakeholders the opportunity to reflect on and contribute to the working definition presented in this Policy Brief. This definition will be evaluated throughout the project and it might change in response to contributions from consortium members or stakeholders in research and innovation practices.

Aside from the working definition and the stakeholders' needs, the workshops focus on collecting promising practices of RRI throughout Europe. These RRI practices will be compiled in an extensive database that is being analysed to (1) formulate good practice standards, (2) select the most promising ones, and (3) make a distinctive set of showcases to present on the **RRI Tools website**. Both the good practice standards and the showcases are meant to guide stakeholders in accomplishing good practice in RRI. The good practice standards, in turn, will contribute to an evaluation methodology of RRI and will be used to build tools for the RRI Toolkit.

Many steps have been taken in realising RRI, but more are necessary. RRI Tools is not the only project active in establishing RRI in Europe. For more in-depth information about the path towards RRI so far and a historical perspective on the development of the concept, we refer to the **About RRI section** on our website; for further information on some of the European projects working on RRI, see below.

OTHER RRI PROJECTS FROM THE EC

ENGAGE2020
The goal of **Engage2020** is to increase the use of engagement methods and policies by mapping what is practiced and spreading awareness of the opportunities amongst researchers, policymakers and other interested parties.

GREAT
The **GREAT** project aims to develop an empirically based and theoretically sound model of the role of responsible research and innovation governance and investigate the characteristics of responsible practices.

PERARES
The **PERARES** project aims to strengthen public engagement in research by involving researchers and Civil Society Organisations in the formulation of research agendas and the research process.

PROGRESS
The **ProGReSS** project aims to advocate a European normative model for RRI globally, using constitutional values as a driver to inform societal desirability.

RESAGORA
The **ResAGorA** project aims at doing extensive research about existing RRI governance across different scientific and technological areas, continuous monitoring of RRI trends and developments in selected countries, and constructive negotiations and deliberation between key stakeholders.

RESPONSIBILITY
The goal of the **Responsibility** project is to develop a virtual observatory for enhancing the interaction among research outcomes and policy making, incorporating the full potential of scientific achievements in the policy development and implementation.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No 612393

RRI Tools