Deliverable 1.3
OBSERVE Horizon Scanning Methodology Report

Project Name	Observing Emergence OBSERVE
Project No.	665136
Project Type	Coordination and Support Action
Project Duration	1.6.2015-30.5.2017 (24 Months)
Project Coordinator	Philine Warnke, Fraunhofer ISI
Funded under	Future and Emerging Technologies FET OPEN CSA
Workpackage	WP 1 Horizon Scanning
Deliverable	3.1: OBSERVE Horizon Scanning Methodology Report
Planned Date	Month 8 (January 2016)
Actual Submission	March 2016
Version	1
Authors	Philine Warnke, Bernd Beckert, Antje Bierwisch, Anselm Geserer, Victoria Kayser, Sven Rung, Elna Schirrmeister; Fraunhofer ISI

Contents

1. Background and Introduction **3**

2 Overall Approach **4**

2.1 How we observe – Horizon Scanning Approach 4

2.2 What we observe 8

3 Implementation **9**

3.1 Scouting Framework 9

3.2 Manual Scouting 12

3.3 Keyword analysis of FET proposals 16

3.4 Web-Mining 48

3.5 Publication Analysis 52

4 Lessons learned/Recommendations **54**

4.1 Scouting Framework 54

4.2 Manual Scouting 54

4.3 Keyword analysis of FET proposals 55

4.4 Webmining 56

4.5 Publication Analysis 57

5 References **57**

1. Background and Introduction

In line with the mission of the European Commission's Future and Emerging Technologies - FET program the overarching aim of OBSERVE is to support Europe to grasp leadership early on in new and emerging technology areas that promise to renew the basis for European competitiveness and growth and that will make a difference for society in the decades to come. For this purpose OBSERVE addresses the concrete objective of identifying new opportunities and directions for interdisciplinary research towards new and visionary technology of any kind.

Next to generating a set of emerging topics fulfilling these criteria through an in depth horizon scanning and multi stakeholder process, OBSERVE aims to support FET in developing a rolling process that can be established by the FET group beyond the project's duration. Furthermore the aim is to spread futures literacy to science and innovation actors within the FET community and Europe as a whole.

Therefore OBSERVE will carefully capture the methodological lessons learned from setting up the OBSERVE radar. This report is specifically dedicated to the four horizon scanning approaches applied in the first phase of the OBSERVE foresight process in order to capture a first set of emerging topics of potential interest for the FET program.

We will document each of the approaches in detail, discuss the experiences and elaborate lessons learned and spell out recommendations for optimization and further development.[1]

1 The methodology applied for the FET Portfolio Analysis is not part of this report as it did not feed the Radar. It can be found in Deliverable 1.1

2 Overall Approach

2.1 How we observe – Horizon Scanning Approach

The OBSERVE Radar can be conceptualised as one element within a strategic process. A strategic process usually evolves cyclically in four phases: strategic intelligence gathering, sense-making, decision making and implementation. OBSERVE is covering the first two phases i.e. strategic intelligence gathering and sense-making which underpin the subsequent phases, decision making and implementation.

As an overarching term for gathering strategic intelligence "**Horizon Scanning**" has recently become widely used. It involves different approaches of information gathering ranging from simple literature analysis to highly sophisticated automated semantic analysis (Amanatidou et al. 2012; Cuhls et al. 2015)

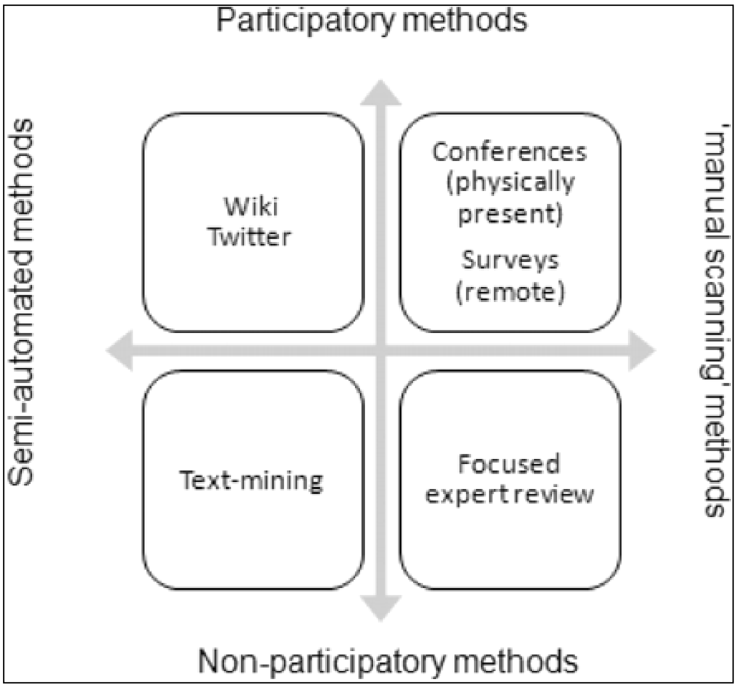


Figure 1: Mapping of Horizon Scanning Methods (Amanatidou et al. 2012, p.212)

For the Horizon Scanning in OBSERVE we have designed a tailored combination of methods each based on its own scientific background that will be explained in more detail below. The main underlying consideration however stems from a set of fundamental insights in the nature of human knowledge and our ability to capture "signals" of change. With (Rossel 2012, p.236) we define: "Weak signals are perceptions of possible changes that are essentially ‘‘candidates’’ (or hypotheses) within a socially relevant and resonant knowledge building process, that in all cases need to be: (1)

- conjugated with other weak or strong signal candidates and iteratively matched against change models (a scenario minded step),
- confronted to one’s own bias-producing capabilities,
- interacted upon with others stakeholders, hopefully involving a diversity of view-points, and
- followed-up and evaluated in light of actual developments, with constant ‘‘early’’ sensitivity regarding strategic options.”

Accordingly, in OBSERVE we are not attempting to “read” signals from the future but to create through our Radar a “socially relevant and resonant knowledge building process” by confronting the biases that are structuring its current perception of the future and to draw attention to “seeds of change” that are currently “discounted” due to the dominant framing of reality of the innovation system (Warnke, Schirrmeister 2016). We feel that such an active, reflexive framing and in particular opening-up for developments at the periphery of society is crucial for identifying breakthrough innovation “that will make a difference for society in the decades to come” as targeted by the FET programme.

Therefore, specific measures to widen the filters and uncover discounted seeds of change are needed and to open up for a diversity of perspectives rather than achieving early consensus (Könnölä et al. 2007). This again requires a sound understanding of the discounting mechanisms in place. Foresight literature provides a number of insights on the different forces that structure our framing. Even though we work with a modified version of Ansoff´s definition of weak signals his categorisation of filters that structure our perception of change remains valid. Ansoff describes three different categories of filters that can lead to biases: surveillance filter, mentality filter and power filter (Ansoff 1975).

The *surveillance filter* structures an organisation’s observation of the environment. Due to resource restriction organisations focus their observation on a few domains where change is expected. Developments outside these domains go unnoticed.

The *mentality filter* is structuring thoughts and judgments of individuals according to their previous experiences. As situations are judged according to the past, those events that do not resonate with past experiences are systematically underestimated. Also groups with similar past experiences such as e.g. decision makers within one organisation or actors within an innovation system may develop their specific mentality filters. In extreme cases such “group thinking” may lead to a systematic discounting of certain domains of change.

Finally, the *power filter* reflects the influence of established hierarchies and power relations, as well as organisational routines on the recognition of certain observations independent of the judgement on its importance.

Other authors have further developed Ansoff’s analysis. For any piece of information to be recognized by an individual or an organization, an active process of seeing and perceiving is required. Neugarten and Day as well as Schoemaker (Day, Schoemaker 2004; Neugarten 2006) use analogies with human vision and distinguish between the focus area and the periphery of an organisational vision. Neugarten (ibid.) stresses the importance of an active vision for any intelligence building process, since for decision-making it is inevitable to discard information: “what we discards depends on what we notice, and what we notice is privileged by being selected as worthy of attention: so by implication we marginalise or ignore the rest” (ibid p.897). Others have analysed biases in foresight studies and have shown that experts tend to judge those developments they think desirable as more likely, in particular when uncertainty is high. This phenomenon has been called “desirability bias” (Ecken et al. 2011) or “overconfidence”. Schoemaker supports these observations and shows in addition biases that are related to the “tendency to look for confirming evidence” for already existing expectations (Schoemaker 2003) and overprediction i.e. the gross overestimation of our capability to predict the future and the respective discard of uncertainty. Finally, Markley has stressed the relevance of credibility when assessing wild cards. He describes different mechanisms leading to low credibility and discounting of wild cards: passive disbelief due to lack of knowledge, active disbelief due to competing convictions, taboos, censorship and disrepute of the “prophet” (Markley 2011, p 1085).

An operationalisation of the findings on “seeds of change” and framing of perception has been developed by (Ilmola, Kuusi 2006). The authors emphasise that restricted capacities to handle information and the trade-off between depth and the width of perception-filters require a continuous decision making process during the search for seeds of change. While width describes the variety of signals and ideas that are taken-up, depth refers to the argumentation effort underpinning the results. They identify four independent variables describing the width and depth of the search including the scope of the briefing, argumentation requirement, number of levels of the analysis and the format of the processing (virtual or intersubjective).

In the light of these background considerations we have designed a methodological framework for the OBSERVE 360° Radar that combines four horizon scanning methods (publication analysis, web mining, portfolio analysis, fringe source scouting) with two sense-making approaches (multi criteria online assessment, creativity workshops).

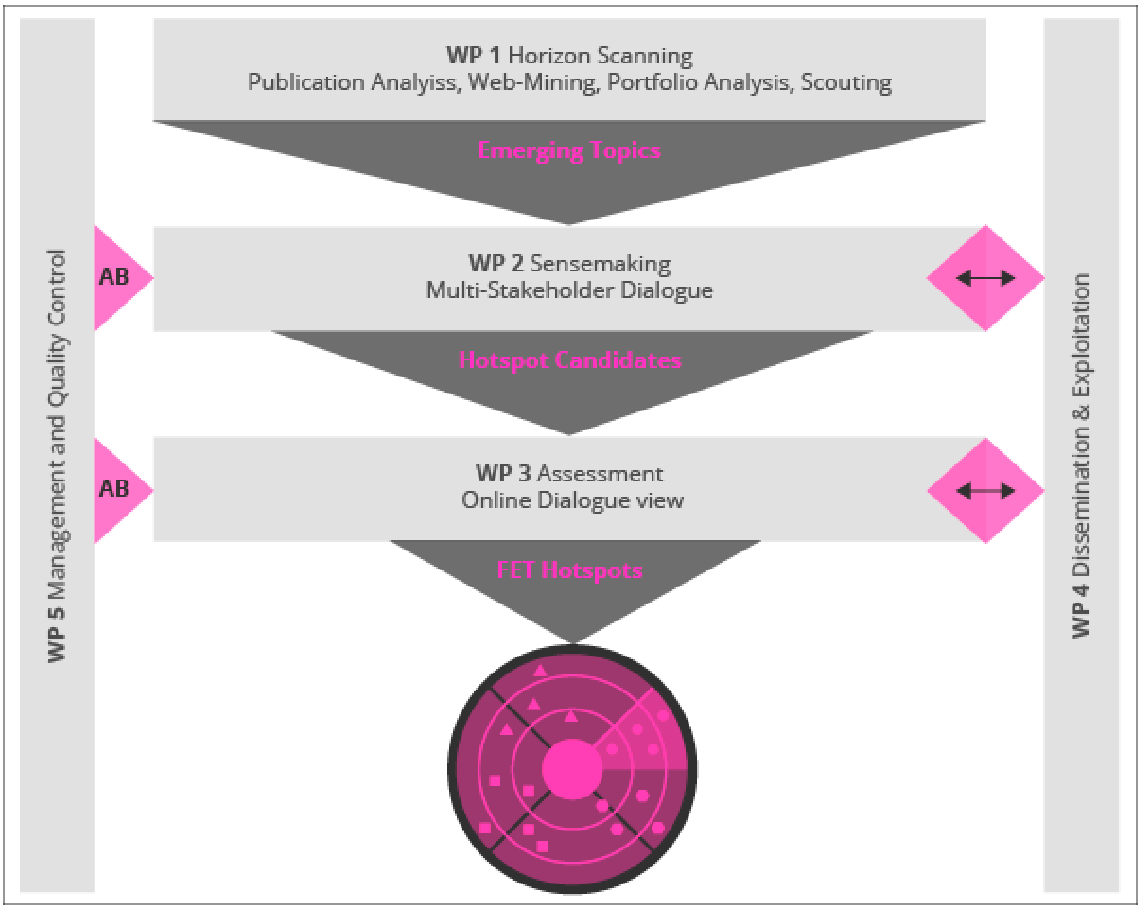


Figure 2: OBSERVE Workflow

The following principles have been observed in order to counteract perception filters and maximise openness for novel and visionary thinking:

Combine rigour and creativity as well as automated analysis and free floating imagination

- Combine online and face to face sense-making processes,
- Systematically recruit sources from the fringes of the mainstream perception,
- Systematically include a diversity of sources from different domains,
- Take utmost care to involve diverse perspectives in terms of cultural context, professional and disciplinary background, gender and age,
- Make use of the power of collective intelligence and human interaction,
- Make extensive use of visual elements to unlock tacit knowledge and imagination,

- Balance wide diversification with strong structuring (one entity carrying out the research with a rigid methodology but full openness to diverse perspectives).

2.2 What we observe

As innovation studies and sociology of technology have long emphasised, relevant impulses for innovation are to be expected not only from science and technology push but also from new practices and needs of users and citizens. In particular demands tend to emerge from the periphery of society and from marginalised groups with special needs (Bijker, Law 1997; von Hippel 2006). This means that serious attempts at identifying novel breakthrough solutions need to go far beyond rough characterisation of "grand challenges of our times" and dive into the dynamics of change at the fringes of society. Therefore, in order to identify FET hotspots it is not sufficient to look at new technologies alone. Rather radical breakthrough innovation is to be expected when novel solutions are aligned with new practices, future needs and new types of collaborations between actors. It is these systemic hotspots where FET can play a role in kick-starting breakthrough innovation with a long-term impact. This is why we place a strong emphasis on S&T issues which are still at the core of FET but complement the findings with a screening of these "soft" aspects.

This is well in line with the recommendations of the FET Advisory Board:

"A key element of the FET mission in the future should be to connect in new ways the creativity of European researchers and the rest of society (citizens, civil society, and other stakeholders). FET should also broaden the definition of innovation beyond technology, to include for example, social sciences, especially with respect to social sciences that result in new challenges and ideas for technology and create innovative collaborations."[2]

In the sense-making phase of OBSERVE these findings will be brought together in order to identify these potential hotspots.

---

2 THE FUTURE OF FET: A possible nucleus for the European Innovation Council. Report from the FET Advisory Group, Brussels, 15th of September 2015; p.7

3 Implementation

3.1 Scouting Framework

Following from the considerations above, the OBSERVE Horizon Scanning was setup to capture the following different types of emerging changes to be fed into the OBSERVE Repository:

- **Solution Idea**

  Emerging technological or social innovation or combination of both addressing a certain problem

- **Science and Technology**

  Novel scientific or technological developments

- **Challenge/Need**

  Emerging challenge or need with potential long term relevance for society

- **Social Practice**

  Emerging change in social practices (new ways of doing) including policy practices

- **Collaboration**

  New formats of collaboration in research and innovation and new constellations of actors collaborating in particular across disciplines.

Combinations of a maximum two categories such as e.g. solution idea/collaboration (i.e. unusual collaboration for a certain solution) were allowed. In the end an additional type **Hybrid** was formed to characterize clusters of findings containing aspects from all types. In order to capture and organize the findings and allow different use cases a database was programmed in Microsoft Access[3].

The following mandatory characteristics were captured for each emerging change hypothesis:

- **Short name (3-10 words)**
- **Short description of the emerging change hypothesis (ca. 5 sentences)**
- **Type of change (max 2, to be selected from the types above)**
- **Source type**
  Selected from: Magazine, Scientific Journal, News Article, Event, FET Proposals, FET Projects, Trend Radar, Tweet, Science Fiction, Aggregation of Scientific Jour-

[3] The full database will be made available to the public later on.

nals (from publication analysis), Artwork, Blog. The source type "Aggregation" was assigned to clusters of findings from different source types.

- **Source:** Here the source was noted that initially brought the topic to the attention of the scouting team.
- **Background material:** Supporting sources often scientific journals referenced in the tweets of news channels.
- **Identification method:** Method that led to the identification of this emerging change hypothesis. Selected from: Web Mining, Scouting Fringe Source, Scouting Antenna, Fringe Interview, Wetzlar Science Fiction Library, Publication Analysis, FET Submission analysis (submitted proposals), FET Portfolio analysis (running projects), Trend Radar.[4]
- **Date of Entry**
- **Found by** (name of the scout first making the entry)
- **Date of publication**
- **Impact Level**

The "impact level" describes the type of impact expected from this development. We distinguish the following four levels:

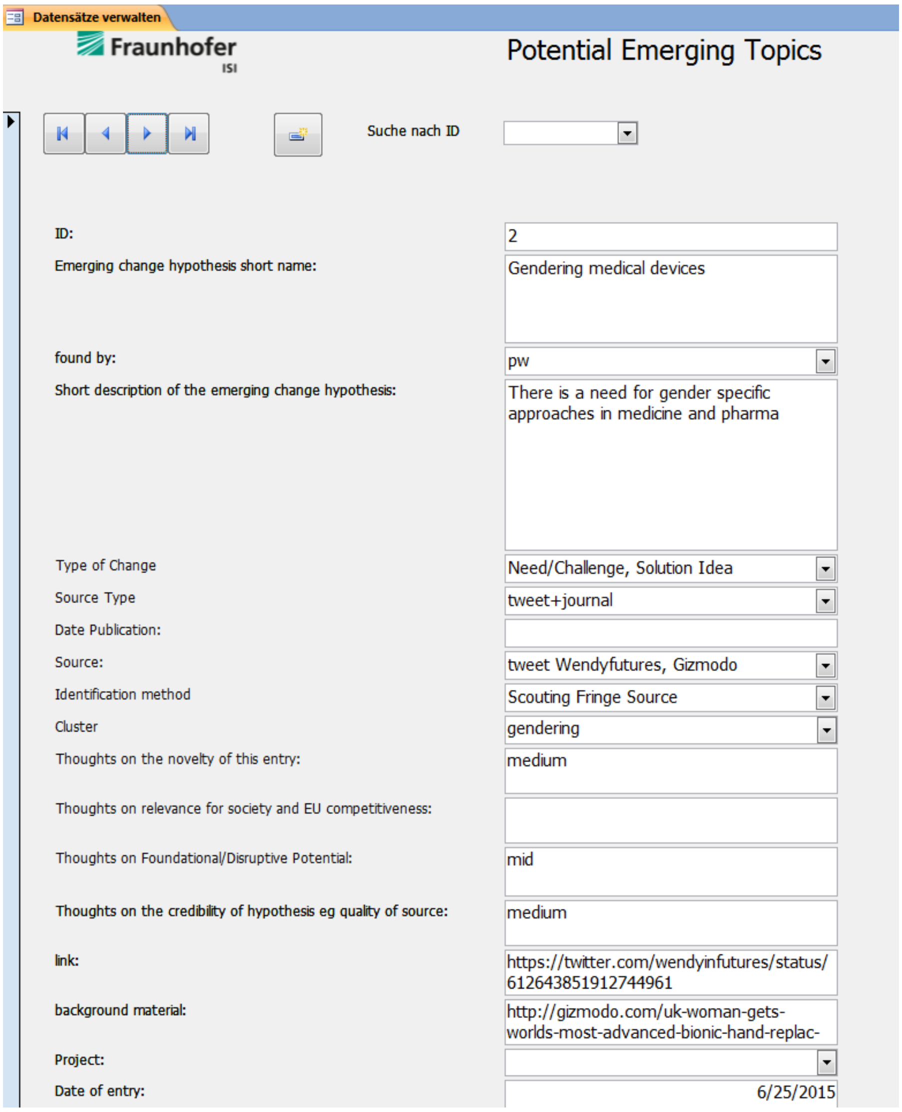
- "Local": impact in a specific domain (e.g. health)
- "Mid Range" impact across several domains e.g. a new materials with applications in health, ICT and energy
- "Widespread" impact across society e.g. fundamental changes in communication patterns
- "Fundamental" impact on the long-term future of the civilisation.

This classification was assigned in a group session among the OBSERVE research team on recommendation of the OBSERVE advisory board. It is by no means an assessment of the relevance or even the relevance for FET. Developments with local impact (e.g. on a certain disease like heart attack) may well be of highest importance.

---

[4] For an explanation of each method see section individual methods below

Figure 1: Microsoft Access Input Mask



Datensätze verwalten

Fraunhofer ISI

Potential Emerging Topics

Suche nach ID

ID:	2
Emerging change hypothesis short name:	Gendering medical devices
found by:	pw
Short description of the emerging change hypothesis:	There is a need for gender specific approaches in medicine and pharma
Type of Change	Need/Challenge, Solution Idea
Source Type	tweet+journal
Date Publication:	
Source:	tweet Wendyfutures, Gizmodo
Identification method	Scouting Fringe Source
Cluster	gendering
Thoughts on the novelty of this entry:	medium
Thoughts on relevance for society and EU competitiveness:	
Thoughts on Foundational/Disruptive Potential:	mid
Thoughts on the credibility of hypothesis eg quality of source:	medium
link:	https://twitter.com/wendyinfutures/status/612643851912744961
background material:	http://gizmodo.com/uk-woman-gets-worlds-most-advanced-bionic-hand-replac-
Project:	
Date of entry:	6/25/2015

Further optional fields served to capture thoughts of the scouts on a preliminary assessment of the findings with respect to novelty, disruptiveness, credibility and relevance. If possible links to online resources were captured.

This basic setup was used to capture findings from all four methods (publication analysis, web mining, FET analysis, manual scouting).

For each approach a team of researchers was working together. Frequent reviews were held to counteract each other's filters. All scouts were encouraged to approach the initial screening with a fully open mindset and to include also seemingly ridiculous findings.

After a screening period of 7 months (July-December 2015) entries were reviewed and identical/very similar ones merged. These cases were specifically marked as they were backed by several rather than only one source. This qualitative aggregation process generated 46 clusters. Seven of these clusters emerged mainly from the webmining. After the scouting phase was finished, the resulting list was subjected to an assessment of relevance by ISI researchers from different realms of who had not been part of the screening. As a consequence a few entries were modified.

After the closing of the scouting process the Horizon Scanning Report (D1.1) and the deck of cards with emerging issues (D4.3) were generated through two specific database queries.

3.2 Manual Scouting

The manual scouting involved a monitoring of communications from diverse sources with a high futures orientation (antennas) and a systematic screening of sources from the fringes of the dominant discourse (fringe sources) as well as an analysis of existing trend repositories. In the manual scouting we adopted a demand pull perspective to complement the science and technology push approach of the publication analysis and web mining. Accordingly, the sources were selected to cover the widest possible range of human demand areas: Mobility, food/nourishment, self-design, health, quality of environment, shelter/housing, personal security, social relations, communication, happiness, self-realisation, meaning, curiosity/learning. The following formats of sources were included: twitter channels, journals/magazines, databases, events, conferences, blogs, news platforms, science fiction literature, futuristic TV series, games, exhibitions, TED talks, crowdfunding platforms.

The tables below gives a list of the sources that were covered for the full length of the scouting period of 8 months as well as conferences, events and Foresight Publications screened manually.

Table 1: Sources monitored in the OBSERVE manual scouting

Twitter Streams
Ars Electronica Futurelab
Arduino/Instructables Twitter
Sociological Review
Institute for the Future IFTF & Futurist Lyn Jeffery
Futurist Wendy Schultz
INSURGE INTELLIGENCE (Fringe Source)
Sharable

The Verge online news platform
Kickstarter (Platform)
Designboom Blog
Nesta, UK
Motherboard online news platform
BBC Futures
Royal Society
The Awl Online Magazine (Fringe Source)
Trends der Zukunft (German Trendradar)
Print Publications
The Economist Intelligent Life (Print Publication)
Lettre International (Print Publication)
Monocle (Print Publication)
The Economist Technology Quarterly (Print Publication)
Events
2nd Barcelona Citizen Science Day, April 13 2015
London Future Fest 2015
CAPS 2015 Networked Social Responsibility
Lift conference 2015 (Shanghai)
Vienna Biennale 2015 Ideas for Change
POC 21 (Paris 08/09 2015)
World Future Society Conference 2014 and 2015 July 24-26
EmTech MIT Nov 2-4 2015
Social Science and Digital Worlds: New Technologies, Old Methods; Wednesday 21st October 2015 University of Leicester
Exhibition at Royal Museums of Fine Arts of Belgium: 2050. A brief history of the future
Globale: Exhibition at ZKM | Center for Art and Media
Foresight Work
Jerome C. Glenn, Elizabeth Florescu, and The Millennium Project Team: 2015-16 State of the Future
100 opportunities for Finland and the world. Radical Technology Inquirer (RTI) for anticipation/evaluation of technological breakthroughs. Finland Committee for the Future 2014
BMBF Foresight Cycle II
Hiltunen, Elina & Hiltunen, Kari. Technolife 2035: How Will Technology Change Our Future? Cambridge 2015
Institute for the Future IFTF 20 Combinatorial Forecasts
Global trends & Future Scenarios (Report Database of the American Development Bank)
WorldWatch Institute VitalSigns
iKnow Weak Weak Signal database

Gartner Top 10 Strategic Technology Trends for 2016 (October 2015)
OECD Horizon Scan of megatrends and technology trends in the context of future research policy. OECD 2016
Futurium Digital Futures. Final Report.

Science Fiction Screening

A specific part of the manual scouting was the screening of science fiction work. This was carried out by the phantastic library in Wetzlar that has collected more than 260.000 future oriented fiction titles performed an in depth screening of recent international science fiction literature from 2005-2015 with a science and technology focus.[5] In total 71 novels and other fictional texts were evaluated and analyzed by the library chairman Thomas Le Blanc. In this context several topics were identified: future of human bodies, networking and surveillance, reality-virtuality amalgamation, changed living environment, economy in the information society, robot’s help and danger of robots, climate fiction and mars literature. Out of this accumulation 12 entries evolved and two of them were placed in clusters.

TED Screening

TED is a non-profit organisation aiming to spread ideas from manifold disciplines – from science to business to global issues - in the form of short talks (18 minutes or less). TED began in 1984 as a conference where Technology, Entertainment and Design converged. This is also how the name TED (Technology, Entertainment, Design) was derived. The annual conference series began in 1990. TED seeks to select emerging artists, scientists and thinkers before they hit the mainstream.[6] Speakers have to be nominated and will be selected from the TED-team. The TED flagship is a one week conference, which is held two times per year. Out of these presentations the „TEDtalks“ are selected and published online on TED.com. Meanwhile, independently run TEDx events are conducted around the world and are also available on TED.com. The TED-talks have been included in the monitoring phase of the OBSERVE project because of the multidisciplinary approach and the claim to spread ideas before they become mainstream, while at the same time demanding scientific standards. Even though TEDx's science guidelines clearly state that information shared from the stage must be supported by peer-reviewed research, this might not me assured in all cases. Therefore only the TEDtalks derived from the original TED conference were included in the moni-

---

5 http://www.phantastik.eu/
Phantastische Bibliothek Wetzlar. Project Future Life

6 https://www.ted.com/about/our-organization/how-ted-works

toring of the OBSERVE project. The monitoring was restricted to talks that were uploaded in 2015 and which were tagged as having a link to technology. The technology-tag was introduced because TEDtalks are not only aiming to inform but also to entertain and therefore popularity of TEDtalks is strongly biased by how appealing and inspiring the presentation is. For the OBSERVE screening talks tagged with “technology” turned out to be best suited.

Kickstarter Platform

Kickstarter is the world's largest funding platform for creative projects and was launched in April 2009.[7] Since its start, more than ten million people have pledged more than $ 2.210.536.933 billion to projects from the world of music, film, art, technology, design, games, fashion, food, publishing and other creative fields.[8] In total more than 100.000 projects were successfully funded in 15 different project categories. The most famous project categories are music (> 22.000 successful funded projects), film (> 19.000 successful funded projects) and publishing (>8.000 successful funded projects).[9]

For the present analysis, the category technology (> 4.000 successful funded projects) was examined in detail to identify interesting projects and possibly thematic priorities that arouse interest in the community or society. For example, the most successful funded projects in the field of technology were[10]:

- Smartwatch Pebble Time[11]
- Pono Music[12]
- Bring Reading Rainbow Back for Every Child, Everywhere[13].

---

7 See http://www.crowdfunding.de/kickstarter/ updated on 17th February 2016.

8 See https://www.kickstarter.com/help/stats?ref=about_subnav updated on 17th February 2016.

9 See https://www.kickstarter.com/help/stats?ref=about_subnav updated on 17th February 2016.

10 See https://www.kickstarter.com/discover/advanced?state=successful&category_id=16&raised=2&sort=most_funded&seed=2426471&page=1 updated on 18th of February 2016.

11 See https://www.kickstarter.com/projects/597507018/pebble-time-awesome-smartwatch-no-compromises updated on 18th February 2016. 78,471 supporters have contributed $ 20,338,986 to enable the realization of the project.

12 See https://www.kickstarter.com/projects/1003614822/ponomusic-where-your-soul-rediscovers-music?ref=discovery updated on 18th February 2016. 18,220 supporters have contributed $ 6,225,354 to facilitate the implementation of the project.

13 See https://www.kickstarter.com/projects/readingrainbow/bring-reading-rainbow-back-for-every-child-everywh?ref=discovery updated on 18th February 2016. 105,857 supporters have contributed $ 5,408,916 to facilitate the implementation of the project.

The analysis has two strands of identification, on the one hand, the analysis of interest in the society to support projects that they view as promising. On the other hand, to uncover project ideas of creative players who are looking for a financing in the community/society. Therefore, projects which were the most popular, most funded or estimated precisely promising, are included in the data collection. The analysis primarily focused on the period end of 2014 to the beginning of 2016. In general, it is to observe that these projects are primarily concerned with the following topics:

- 3D printing, its applications and possible new materials,
- the use of certain materials such as carbon fiber in different application fields (cross-innovation),
- DIY electronic projects,
- new and efficient ways of energy storage,
- space exploration, participation or ownership of the society in space.

As a final outcome the OBSERVE database contains single 91 entries exclusively relying on the manual scouting. Of these 9 emerged from previous Foresight work and 7 from science fiction analysis. In addition the majority of the aggregated clusters of findings originated mainly from the manual scouting.

3.3 Keyword analysis of FET proposals

This section describes the method used to identify new and emerging research topics based on the keyword analysis of the FET proposals-database. The result of the keyword analysis is a list and short descriptions of 19 topics which were fed into the Horizon Scanning database of observe.

The assumption of this exercise is that FET Open and FET Proactive projects are indicators for new topics in research and technology. Both schemes encourage researchers to turn in research proposals that explore new ideas, concepts, approaches and technologies in a collaborative and trans-disciplinary manner. There is a special emphasis on non-mainstream and high-risk research.

In a separate deliverable (D1.1 FET Portfolio Report) we have used a bottom-up topical clustering approach of recent FET Open projects to identify new and emerging research topics. This bottom-up clustering was based on relatively small sample of 67 recent FET Open projects allowing for a manual analysis.

The keyword analysis is a different approach to find patterns and clusters in the FET sample. It uses a semi-automated method and becomes possible because of the high number of cases (more than 4.000) which are research project proposals in our case.

For the keyword analysis we use a database provided by the European Commission which contains acronyms, abstracts and keywords of research proposals for FET Open and FET Proactive projects. The database contains 2.880 project proposals turned in during FP7 and 1.146 proposals turned in during the first rounds of H2020. The sample can be characterized as follows:

- 2.889 FP7 FET Open and FET Proactive proposals including the set of funded projects. In the FP7 set of the sample, freely given keywords by coordinators are available.
- 1.146 H2020 FET Open and FET Proactive proposals including the set of funded projects. In the H2020 set of the sample, pre-given keywords are available as well as freely given keywords by the coordinators.

For a first analysis, the two datasets were merged so that the total number of proposals adds up to 4.035.

In the following, we will shortly document two approaches which yielded interesting results but which were not suited to deliver a robust list of new and emerging topics for the observe Horizon Scanning exercise (Step 1 and 2). In more detail we will then document the method which lead to the final list of topics and which consists of an automated counting and a manual clustering (Step 3).

Step 1: Counting the words „new“, „first“, „unconventional“, and „innovative“ in proposals´ abstracts.

In a first attempt to identify proposals with a high level of novelty we searched for the words „new“, „first“, „unconventional“, and „innovative“ in proposals´ abstracts. We assumed that the use of several of these words in the abstracts reflects the assessments of the researchers that their proposal will explore new and uncharted terrain.

We have found 38 proposals which have used three of the four keywords. For example, the proposal for a project with the acronym AstroBYTE used the word “new” three times, the word “first” once and the word “innovative” once:

Abstract of the research proposal AstroBYTE:

"There is a pressing need for hardware engineers to explore new approaches to fault detection, fault diagnosis and self-repair, as current strategies are constrained with a reliance on a central controller, and architectural constraints are placed on the number of faults (e.g. open/short-circuits) that can be tolerated, and on the level of granularity at which repairs can be implemented. AstroBYTE is an ambitious proof-of-concept project that exploits the role of astrocytes in neurological self-repairing systems to deliver a radically new, highly robust and adaptive information processing computing architecture. AstroBYTE will perform experiments to capture data on how astrocytes interact with neurons and perform distributed fault detection and repair. A computational model will be derived from data to realise the self-repairing mechanisms in software. The core innovative step is the understanding and modelling of the repair coordination mechanisms. A pathway to the FPGA-based hardware implementation of the self-repairing mechanism will be established, with the first-ever hardware building blocks which merge astrocytes and spiking neurons. Two hardware demonstrators will be developed (1) Image filtering of visual sensor (retina) data for object detection and, (2) Multi-motor controller for robot positioning, which will showcase how this novel brain based self-repair capability can overcome the limitations of existing repair approaches. In particular, online access to the hardware demonstrators will be implemented via a remote FPGA lab for registered researchers, enabling demonstrator execution and exploration of other applications.

In summary, AstroBYTE will establish the foundation for a new direction of research in highly robust self-repairing neural-based computing systems. In the long term, AstroBYTE will provide a flexible, accessible research platform to allow neuroscientists to explore the role of astrocytes in brain function."

Table 1 displays the list of proposals which have used at least three of the four keywords.

Table 1: List of proposals using in their abstract the keywords „new" „first", „unconventional" and/or „innovative"

ALIGN
AstroBYTE
BLASE
BS4ICTRSRCH

ChipInCell
CLEVER
CREATION
DBLOGIC
DIGENESIS
ECNC
ECPH
E-LAB
ELECTRANSNANO
EROS
FACE
GRASP
HIDO
I2Mi
i-RISC
LC-NaM
MATISSE
MOLIONS
Mutual e-Currency
NECTAR
NewDevices
NFG-Photonics

NICK-I
NUT
OPTI4PLM
PAACOmm
PIEPER
PROMISCE
QUANTOOL
REMOSIGN
SHERPA
UPGRADE
USE-NOW
WINE-CELLAR

Of these 38 research proposals, six (or 16%) were finally funded (BS4ICTRSRCH, GRASP, HIDO, i-RISC, PROMISCE, and UPGRADE), an acceptance rate which is only slightly higher than the overall acceptance rates which we estimate to be around 10 percent for FP7 proposals and 5 percent for H2020 proposals.[14] Thus, this list does not provide a solid base for identifying new research topics. What is more, this approach takes the announcements of the proposers at face value which in some cases might not be justified. The fact that proposers claim that they will explore new, innovative or unconventional topics or apply such methods does not necessarily mean that this is the case. Another interpretation could be that they have only very well understood what the programme intentions are and what reviewers will honor. The rhetoric in the abstract surely does not substitute for a thorough examination of the actual idea. However, as we wanted to concentrate on automated methods for identifying new topics we did not proceed here but decided to try another approach.

14 We manually checked, how many of the proposals in the database were actually funded and calculated the numbers mentioned. However, this is just an estimation because of time lags between turning in a proposal and the actual start of the project.

Step 2: Frequency counting of pre-given keywords in the H2020-share of the sample

In the H2020-share of the database, proposers provided keywords characterizing their research project from a list of pre-given keywords. We tried an analysis of these keywords and found the following frequencies (see table 2: Frequency of pre-given keywords in H2020 proposals)

Table 2: Frequency of pre-given keywords in H2020 proposals

Pre-given keywords	**Frequency**
Atomic, molecular physics	31
Diagnostic tools (e.g. genetic, imaging)	31
Artificial Intelligence (AI)	30
Electronic properties of materials and transport	30
Artificial intelligence, intelligent systems, mult	27
Electrical and electronic engineering: semiconduct	24
Algorithms and complexity	19
Cognitive science, human computer interaction, nat	17
Electrochemistry, electrodialysis, microfluidics,	17
Application of mathematics in sciences	16
Automation, Robotics Control Systems	16
Advanced Systems Architecture	14
Biophysics	14
Electronic properties of materials, surfaces, inte	14
Communication networks, media, information society	13
Condensed matter physics (including formerly solid	13
Biological systems analysis, modelling and simulat	12

Biomaterials synthesis	12
Chemical reactions: mechanisms, dynamics, kinetics	12
Electronics, photonics	12
Diagnostic and implantable devices, environmental	11
Lasers, ultra-short lasers and laser physics	11
Bioinformatics	10
Environment, resources and sustainability	10
Materials for sensors	10

The result of this counting was in fact disappointing because the pre-given keywords denote disciplines or research fields rather than new developments or topics. Also, there is an obvious similarity between the pre-given keywords and the twelve topical clusters which the European Commission provides to characterize the portfolio of its FET programme (AI & Cognition, Complexity, Bio- & Neuro-ICT, etc., see http://ec.europa.eu/digital-agenda/en/fet-projects-portfolio).

Furthermore, it seems that the classification is not really selective. For example the pre-given keyword combination “Artificial Intelligence (AI)” overlaps with the combination “Artificial intelligence, intelligent systems, mult”. It is unclear how to explain this overlap and how to proceed with such entries. Thus, we decided for another approach.

Step 3: Analysis of freely given keywords in a two-step clustering exercise

In contrast to the pre-given keywords which were available only for the newer H2020 proposals, freely given keywords were available for both parts of the sample, e.g. for H2020 as well as for FP7 proposals. Altogether the dataset provided freely given keywords for 4.035 proposals.

Some proposers characterized their idea extensively using up to 15 keywords, some proposers were rather short and provided only two or three keywords or keyword combinations.

In the keyword-fields there were several typing errors, and topics were sometimes given in singular, sometimes in the plural form. These had to be corrected manually, all else was calculated automatically. The result of the counting is displayed in table 3:

Table 3: Frequencies of freely given keywords in the complete dataset

Free keywords given by proposers (FP7 and H2020 dataset, N=4.035)	**Frequency of occur-ance**
Machine Learning	114
Complex system	74
Nanotechnology	60
Data mining	58
Neuroscience	57
Quantum dot	57
Microfluid	54
Human-Computer Interaction	50
Plasmonics	50
artificial intelligence	49
EEG	41
Quantum computing	41
social network	38
Ontologies	37
Spintronics	37
Computer Vision	36
Nanostructures	36
virtual reality	36
Metamaterials	35
CMOS	34
Robotics	34
Quantum Technology	33

Real-time	32
carbon nanotubes / structures	31
Photonic crystal	31
Quantum Optics	31
Cloud Computing	30
Complex network	30
computational model	30
Quantum communication	30
Biosensor / sensing	28
Nanoparticles	28
Quantum information processing	28
Terahertz	28
Distributed Systems	27
Nanoelectronics	26
Self-organization	26
MEMS	25
Wireless Sensor Networks	25
molecular electronics	24
Nanophotonics	24
quantum simulation	24
Signal Processing	24
silicon photonics	24
energy harvesting	22
Nonlinear Optics	22
Systems Biology	22
Augmented Reality	21
Decision Support System	21
Nanowires	21
Compressed Sensing	20

Pervasive Computing	20
Sensor network	20
Trust	20
FMRI	19
High Performance Computing	19
Knowledge Representation	19
Medical Imaging	19
pattern recognition	19
wireless communications	19
Cognitive Systems	18
natural language processing	18
Photovoltaic	18
Semantic Web	18
Semiconductors	18
Software engineering	18
Verification	18
Bioinformatics	17
Cognitive Radio	17
Fault tolerant computing	17
Information Retrieval	17
Optical communications	17
Spectroscopy	17
Superconductivity	17
synthetic biology	16
ubiquitous computing	16
Crowdsourcing	15
Cyber Physical Systems	15
Distributed Computing	15
Lab-on-a-Chip	15

Optoelectronics	15
Sustainability	15
data analysis	14
Decoherence	14
integrated optics	14
Internet of Things	14
Microelectronics	14
multi-agent systems	14
NEMS	14
Embodiment	13
linked data	13
RFID	13
Self-assembly	13
Visual Analytics	13
Adaptive Systems	12
embodied cognition	12
GaN	12
social computing	12
social media	12
Tissue engineering	12
unconventional computing	12
User Interfaces	12
agent-based simulation / modeling	11
Computational Fluid Dynamics	11
Future Internet	11
GaN	11
Open Source	11
Optogenetics	11
Virtualization	11

human-robot interaction	10
knowledge discovery	10
magnetic nano	10
Neuroimaging	10
parallel computing	10
reinforcement learning	10
renewable energy	10
smart materials	10
Swarm Intelligence	10
Web 2.0	10
(adaptive) antennas	9
Biomarkers	9
evolutionary computation	9
frequency comb	9
Optomechanics	9
photonic integrated circuits	9
reconfigurable computing	9
remote sensing	9
serious games	9
Autonomous Systems	8
computational biology	8
Computational linguistics	8
computational neuroscience	8
context awareness	8
environmental monitoring	8
HPC	8
Magnetic Resonance Imaging	8
MEG	8
mobile devices	8

Neurophysiology	8
Privacy Protection	8
self-searching	8
Smart textiles	8
Social Signal Processing	8
Solar cells	8
speech recognition	8
text mining	8
Topological Quantum Computation	8
trapped ions	8
Ultracold atoms	8
User Experience	8
affective computing	7
autonomic computing	7
Biomimetics	7
Collective Behaviour	7
collective intelligence	7
epitaxial growth	7
Exoskeleton	7
flexible electronics	7
FPGA	7
Functional Programming	7
HCI	7
multimodal interfaces	7
Neuroaesthetics	7
Neuroengineering	7
spinal cord injury	7
Surface Acoustic Wave	7
Telecommunication	7

Usability	7
visual perception	7
wearable sensors	7
3D Integration	6
AFM	6
BCI	6
Big data Analytics	6
Cold Atoms	6
Drug Discovery	6
dynamical systems	6
Haptic interfaces / displays	6
mobile services	6
Multimodal Interaction	6
Neurorehabilitation	6
prediction model	6
smart cities	6
social web	6
Software development	6
swarm robotics	6
topological insulators	6
supramolecular chemistry	5
UAV	5
VLSI	5

What the numbers mean:

“Machine learning: 114” means that in 114 proposals the free keyword „Machine learning“ was given by proposers.

We also checked for multiple classifications, for example when a proposal was characterized by “Machine learning” as well as “Complex system”, but found that this is not very often the case.

The automatically generated results are quite interesting, however, we found the list to be too long. It is obvious that there are topical overlaps in the list of freely given keywords. What used to be a unique combination to characterize a research idea was taken apart in order to count frequencies. In a second step, the separated keywords needed to be clustered again into overarching topics. This step was done manually by experts.

We started the clustering exercise with the top 10 keywords and assigned keywords following later in the list, according to similarity. The leading question was: “Do these research fields fit together, are they related in any form?” If yes, we allocated the respective keyword to the bigger cluster. After all top 10 keyword entries were complemented in this way we continued with the remaining entries and allocated related keywords in the same fashion. The reallocation process was guided by external characterizations of the research fields which were compiled from Web-sources, mostly from Wikipedia.

In a preceding search we tested the suitability of Wikipedia as a source for our exercise and found it very reliable and accurate compared to field descriptions found elsewhere. For some characterizations we involved Fraunhofer experts, especially in the areas of biotechnology and nanotechnology. The result of our clustering exercise is displayed in table 4.

Table 4: Clustering the freely-given keywords into 19 topics

Free keywords given by proposers	Frequency of occurance	What is it about?
Machine Learning	444 (114) (in brackets: occurance of keyword alone, no brackets: whole cluster occurence)	Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions, rather than following strictly static program instructions. Machine learning is closely related to computational statistics; a discipline that aims at the design of algorithm for implementing statistical methods on computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms is infeasible. Example applications include spam filtering, optical character recogni-

		tion (OCR), search engines and computer vision. Machine learning is sometimes conflated with data mining, although that focuses more on exploratory data analysis. Machine learning and pattern recognition can be viewed as two facets of the same field. (Source: Wikipedia “machine learning”) Closely connected with: data mining (58), computer vision (36), computational model (30), medical imaging (19), pattern recognition (19), data analysis (14), visual analytics (13), speech recognition (8), text mining (8), big data analytics (6), prediction model (6), software development (6), natural language processing (18), agent-based simulation/modeling (11), multi-agent systems (14), signal processing (24), compressed sensing (20), reinforcement learning (10), knowledge discovery (10). Added associations: 330
Nano-technology	341 (60)	Nanotechnology is manipulation of matter on an atomic, molecular, and supramolecular scale. A description of nanotechnology was established by the National Nanotechnology Initiative, which defines nanotechnology as the manipulation of matter with at least one dimension sized from 1 to 100 nanometers. This definition reflects the fact that quantum mechanical effects are important at this quantum-realm scale, and so the definition shifted from a particular technological goal to a research category inclusive of all types of research and technologies that deal with the special properties of matter that occur below the given size threshold. It is therefore common to see the plural form "nanotechnologies" as well as "nanoscale technologies" to refer to the broad range of research and applications whose common trait is size. Because of the variety of potential applications (including industrial and military), governments have invested billions of dollars in nanotechnology research. Nanotechnology as defined by size is naturally very broad, including fields of science as diverse as surface science, organic chemistry, molecular biology, semiconductor physics, microfabrication, etc. The associated research and applications are equally diverse, ranging from extensions of conventional device physics to completely new approaches based upon molecular self-assembly, from developing new materials with dimensions on the nanoscale to direct control of matter on the atomic scale. (Source: Wikipedia “nano-technology”) Closely connected with: Nanostructures (36), carbon nanotubes/structures (31), nanoparticles (28), nanoelectronics (26), molecular electronics (24), nanophotonics (24), nanowires (21), magnetic nano (10), epitaxial growth (7), NEMS (nanoelectromechanical systems) (14), MEMS (micro-electro-mechanical systems) (25), metamaterials (35) Added associations: 281
Quantum dot	311 (57)	A Quantum dot (QD) is a crystal of semiconductor material whose diameter is on the order of several nanometers - a size which results in its free charge carriers experiencing "quantum confinement" in all three spatial dimensions. The electronic properties of quantum dots are intermediate between those of bulk semiconductors and of discrete molecules and closely related to their size and shape. This allows properties such as the band gap, emission color, and absorption

		spectrum to be highly tuneable, as the size distribution of quantum dots can be controlled during fabrication. For example, the band gap in a quantum dot, which determines the frequency range of emitted light, is inversely related to its size. In fluorescent dye applications, the frequency of emitted light increases as the size of the quantum dot decreases, shifting the color of emitted light from red to violet. Researchers have studied applications for quantum dots in transistors, solar cells, LEDs, and diode lasers. They have also investigated quantum dots as agents for medical imaging and as possible qubits in quantum computing. The small size of quantum dots allows them to be suspended in various solvents and thus compatible with solution processing techniques such as spin coating and inkjet printing. These are inexpensive compared to conventional semiconductor device fabrication involving small areas and ultra-high vacuum. The first commercial release of a product utilizing quantum dots was the Sony XBR X900A series of flat panel televisions released in 2013. (Source: Wikipedia “quantum dot”) Closely connected with: Quantum computing (41), quantum technology (33), quantum optics (31), quantum communication (30), quantum information processing (28), quantum simulation (24), topological quantum computation (8), Trapped ions (8), ultracold atoms (8), cold atoms (6), superconductivity (17), topological insulators (6), decoherence (14) Added associations: 254
Complex system*	266 (74)	Complex systems present problems both in mathematical modelling and philosophical foundations. The study of complex systems represents a new approach to science that investigates how relationships between parts give rise to the collective behaviors of a system and how the system interacts and forms relationships with its environment. The equations from which models of complex systems are developed generally derive from statistical physics, information theory and non-linear dynamics and represent organized but unpredictable behaviors of natural systems that are considered fundamentally complex. The physical manifestations of such systems are difficult to define, so a common choice is to identify "the system" with the mathematical information model rather than referring to the undefined physical subject the model represents. Such systems are used to model processes in computer science, biology,economics, physics, chemistry and many other fields. It is also called complex systems theory, complexity science, study of complex systems, sciences of complexity, non-equilibrium physics, and historical physics. A variety of abstract theoretical complex systems is studied as a field of mathematics. The key problems of complex systems are difficulties with their formal modelling and simulation. From such a perspective, in different research contexts complex systems are defined on the basis of their different attributes. Since all complex systems have many interconnected components, the science of networks and network theory are important aspects of the study of complex systems. A consensus regarding a single universal definition of complex system does not yet exist. For systems that are less usefully represented with equations various other kinds of narratives and methods for identifying, exploring, de-

	signing and interacting with complex systems are used. (Source: Wikipedia “complex systems”) Closely connected with: Complex networks (30), distributed systems (27), self-organisation (26), nonlinear optics (22), systems biology (22), adaptive systems (12), swarm intelligence (10), evolutionary computation (9), affective computing (7), autonomic computing (7), collective behavior (7), collective intelligence (7), dynamical systems (6), swarm robotics (6) Added associations: 192

Social network	230 (38)	A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a set of the dyadic ties between these actors. The social network perspective provides a set of methods for analyzing the structure of whole social entities as well as a variety of theories explaining the patterns observed in these structures. The study of these structures uses social network analysis to identify local and global patterns, locate influential entities, and examine network dynamics. Social networks and the analysis of them is an inherently interdisciplinary academic field which emerged from social psychology, sociology, statistics, and graph theory. Social network analysis is now one of the major paradigms in contemporary sociology, and is also employed in a number of other social and formal sciences. Together with other complex networks, it forms part of the nascent field of network science. (source: Wikipedia “social network)” Closely connected with: Decision support system (21), Semantic web (18), crowdsourcing (15), social computing (12), social media (12), web 2.0 (10), serious games (9), privacy protection (8), user experience (8), usability (7), smart cities (6), social web (6), decision support system (21), open source (11), information retrieval (17), linked data (13), virtualization (11) Added associations: 192 Partly overlaps to robotics (for example swarm intelligence)
Neuro-science	209 (57)	Neuroscience is the scientific study of the nervous system. Traditionally, neuroscience has been seen as a branch of biology. However, it is currently an interdisciplinary science that collaborates with other fields such as chemistry, cognitive science, computer science, engineering, linguistics, mathematics, medicine (including neurology), genetics, and allied disciplines including philosophy, physics, and psychology. It also exerts influence on other fields, such as neuroeducation, neuroethics, and neurolaw. The term neurobiology is usually used interchangeably with the term neuroscience, although the former refers specifically to the biology of the nervous system, whereas the latter refers to the entire science of the nervous system. The scope of neuroscience has broadened to include different approaches used to study the molecular, cellular, developmental, structural, functional, evolutionary, computational, and medical aspects of the nervous system. The techniques used by neuroscientists have also expanded enormously, from molecular and cellular studies of individual nerve cells to imaging of sensory and motor tasks in the brain. Recent theoretical advances in neuroscience have also been aided by the study of neural networks. (Source: Wikipedia “neuroscience”) Closely connected with: EEG (41), Neuroimaging (10), computational neuroscience (10), neurophysiology (8), neuroaesthetics (7), neuroengineering (7), spinal cord injury (7), neurorehabilitation (6), BCI (6), optogenetics (11), magnetic resonance imaging (8), FMRI (19), computational neuroscience (8), MEG (8), neurophysiology (8), drug discovery (6). Added associations: 152 (Within the Neuro-cluster, the single-most important field is EEG)

Cloud Computing	199 (30)	Cloud computing, also known as on-demand computing, is a kind of Internet-based computing, where shared resources and information are provided to computers and other devices on-demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in third-party data centers. It relies on sharing of resources to achieve coherence and economies of scale, similar to a utility (like the electricity grid) over a network. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. (source: Wikipedia “cloud computing”). Closely connected with: Pervasive computing (20), high performance computing (19), software engineering (18), fault tolerant computing (17), ubiquitous computing (16), distributed computing (15), parallel computing (10), reconfigurable computing (9), verification (18), unconventional computing (12), HPC (high-performance computing) (8), functional programming (7) Added associations: 169
Biosensor / sensing	192 (28)	A biosensor is an analytical device, used for the detection of an analyte, that combines a biological component with a physicochemical detector. It consists of three elements: First the sensitive biological element (e.g. tissue, microorganisms, organelles, cell receptors, enzymes, antibodies, nucleic acids, etc.), a biologically derived material or biomimetic component that interacts (binds or recognizes) the analyte under study. Second the transducer or the detector element (works in a physicochemical way; optical, piezoelectric, electrochemical, etc.) that transforms the signal resulting from the interaction of the analyte with the biological element into another signal (i.e., transduces) that can be more easily measured and quantified. Third, the biosensor reader device with the associated electronics or signal processors that are primarily responsible for the display of the results in a user-friendly way. (source: Wikipedia “biosensor”) Closely connected with: Wireless Sensor networks (25), sensor network (20), lab-on-a-chip (15), RFID (13), Smart materials (10), biomarkers (9), remote sensing (9), environmental monitoring (8), smart textiles (8), wearable sensors (7), Cyber Physical Systems (15), Internet of Things (14), future Internet (11) Added associations: 164 Although different technologies are being used (biology, computer science, communication technology), projects have the similarity that they are developing sensors in one way or the other.
Human-Computer Interaction	152 (50)	Human–computer interaction (HCI) researches the design and use of computer technology, focusing particularly on the interfaces between people (users) and computers. Researchers in the field of HCI both observe the ways in which humans interact with computers and design technologies that let humans interact with computers in novel ways. As a field of research, Human-Computer Interaction is situated at the intersection of computer science, behavioral sciences, design, media studies, and several other fields of study. The term connotes that, unlike other tools

		with only limited uses (such as a hammer, useful for driving nails, but not much else), a computer has many uses and this takes place as an open-ended dialog between the user and the computer. The notion of dialog likens human-computer interaction to human-to-human interaction, an analogy the discussion of which is crucial to theoretical considerations in the field. (Source: Wikipedia “human computer interaction”) Closely connected with: virtual reality (36), augmented reality (21), user interfaces (12), HCI (7), multimodal interfaces (7), visual perception (7), haptic interfaces / displays (6), multimodal interaction (6) Added associations: 102 (thematic link to robotics-Cluster)
artificial intelligence	131 (49)	Artificial intelligence (AI) is the intelligence exhibited by machines or software. It is also the name of the academic field of study which studies how to create computers and computer software that are capable of intelligent behavior. Major AI researchers and textbooks define this field as "the study and design of intelligent agents”, in which an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success. John McCarthy, who coined the term in 1955, defines it as "the science and engineering of making intelligent machines". AI research is highly technical and specialized, and is deeply divided into subfields that often fail to communicate with each other. Some of the division is due to social and cultural factors: subfields have grown up around particular institutions and the work of individual researchers. AI research is also divided by several technical issues. Some subfields focus on the solution of specific problems. Others focus on one of several possible approaches or on the use of a particular tool or towards the accomplishment of particular applications. The central problems (or goals) of AI research include reasoning, knowledge, planning, learning, natural language processing (communication), perception and the ability to move and manipulate objects. General intelligence is still among the field's long-term goals. Currently popular approaches include statistical methods, computational intelligence and traditional symbolic AI. There are a large number of tools used in AI, including versions of search and mathematical optimization, logic, methods based on probability and economics, and many others. The AI field is interdisciplinary, in which a number of sciences and professions converge, including computer science, mathematics, psychology, linguistics, philosophy and neuroscience, as well as other specialized fields such as artificial psychology. (Source: Wikipedia “artificial intelligence”). Closely connected with: Ontologies (37), knowledge representation (19), computational linguistics (8), cognitive systems (18) Added associations: 82 Comment: Initially considered too broad a description and interdisciplinary by definition. Researchers who have given this keyword have just flagged the broad field but not the subfield which they put their subject in.

Photonic crystal	131 (31)	A photonic crystal is a periodic optical nanostructure that affects the motion of photons in much the same way that ionic lattices affect electrons in solids. Photonic crystals occur in nature in the form of structural coloration—and, in different forms, promise to be useful in a range of applications. Photonic crystals can be fabricated for one, two, or three dimensions. One-dimensional photonic crystals can be made of layers deposited or stuck together. Two-dimensional ones can be made by photolithography, or by drilling holes in a suitable substrate. Fabrication methods for three-dimensional ones include drilling under different angles, stacking multiple 2-D layers on top of each other, direct laser writing, or, for example, instigating self-assembly of spheres in a matrix and dissolving the spheres. Photonic crystals can, in principle, find uses wherever light must be manipulated. Existing applications include thin-film optics with coatings for lenses. Two-dimensional photonic-crystal fibers are used in nonlinear devices and to guide exotic wavelengths. Three-dimensional crystals may one day be used in optical computers. (source: Wikipedia “photonic crystal”) Closely connected with: silicon photonics (24), spectroscopy (17), optoelectronics (15), integrated optics (14), optomechanics (9), photonic integrated circuits (9), GaN (12) Added associations: 100
Terahertz	106 (28)	In May 2012, a team of researchers from the Tokyo Institute of Technology published in Electronics Letters that it had set a new record for wireless data transmission by using Terahertz-rays and proposed they be used as bandwidth for data transmission in the future. The team's proof of concept device used a resonant tunneling diode (RTD) in which the voltage decreased as the current increased, causing the diode to "resonate" and produce waves in the terahertz band. With this RTD, the researchers sent a signal at 542 GHz, resulting in a data transfer rate of 3 Gigabits per second. The demonstration was twenty times faster than the current Wi-Fi standard and doubled the record for data transmission set the previous November. Potential uses exist in high-altitude telecommunications, above altitudes where water vapor causes signal absorption: aircraft to satellite, or satellite to satellite. (source: Wikipedia “terahertz”) Closely connected with: wireless communication (19), optical communications (17), adaptive antennas (9), frequency comb (9), telecommunication (7), cognitive radio (17) Added associations: 78

Robotics	105 (34)	Robotics is the branch of mechanical engineering, electrical engineering, electronic engineering and computer science that deals with the design, construction, operation, and application of robots, as well as computer systems for their control, sensory feedback, and information processing. These technologies deal with automated machines that can take the place of humans in dangerous environments or manufacturing processes, or resemble humans in appearance, behavior, and/or cognition. Many of to-day's robots are inspired by nature contributing to the field of bio-inspired robotics. (source: Wikipedia "robotics") Closely connected with: embodiment (13), self-assembly (13), embodied cognition (12), human-robot-interaction (10), autonomous systems (8), context awareness (8), exoskeleton (7) Added associations: 71
CMOS	98 (34)	Complementary metal–oxide–semiconductor (CMOS) is a technology for constructing integrated circuits. CMOS technology is used in microproces-sors, microcontrollers, static RAM, and other digital logic circuits. CMOS technology is also used for several analog circuits such as image sensors (CMOS sensor), data converters, and highly integrated transceivers for many types of communication (source: Wikipedia "CMOS"). Closely connected with: Semiconductors (18), microelectronics (14), flexi-ble electronics (7), FPGA (7), surface acoustic wave (7), 3D integration (6), VLSI (5) Added associations: 64
Microfluid*	65 (54)	Microfluidics is a multidisciplinary field intersecting engineering, physics, chemistry, biochemistry, nanotechnology, and biotechnology, with practical applications to the design of systems in which low volumes of fluids are processed to achieve multiplexing, automation, and high-throughput screening. Microfluidics emerged in the beginning of the 1980s and is used in the development of inkjet printheads, DNA chips, lab-on-a-chip technol-ogy, micro-propulsion, and micro-thermal technologies. It deals with the behavior, precise control and manipulation of fluids that are geometrically constrained to a small, typically sub-millimeter, scale. Typically, micro means one of the following features: small volumes (µL, nL, pL, fL), small size, low energy consumption, effects of the micro domain. Typically fluids are moved, mixed, separated or otherwise processed. Nu-merous applications employ passive fluid control techniques like capillary forces. In some applications external actuation means are additionally used for a directed transport of the media. Examples are rotary drives applying centrifugal forces for the fluid transport on the passive chips. Active micro-fluidics refers to the defined manipulation of the working fluid by active (micro) components such as micropumps or micro valves. Micro pumps supply fluids in a continuous manner or are used for dosing. Micro valves determine the flow direction or the mode of movement of pumped liquids. Often processes which are normally carried out in a lab are miniaturized on a single chip in order to enhance efficiency and mobility as well as reducing sample and reagent volumes. (Source: Wikipedia "microfluidics") Closely connected with: Computational fluid dynamics (11) Added associations: 11 (very few associations to be listed)

Fraunhofer ISI

Energy harvesting	58 (22)	Energy harvesting is the process by which energy is derived from external sources (e.g. solar power, thermal energy, wind energy, salinity gradients, and kinetic energy), captured, and stored for small, wireless autonomous devices, like those used in wearable electronics and wireless sensor networks. Energy harvesters provide a very small amount of power for low-energy electronics. While the input fuel to some large-scale generation costs resources (oil, coal, etc.), the energy source for energy harvesters is present as ambient background and is free. For example, temperature gradients exist from the operation of a combustion engine and in urban areas, there is a large amount of electromagnetic energy in the environment because of radio and television broadcasting. (source: Wikipedia “energy harvesting”) Closely connected with: photovoltaic (18), renewable energy (10), solar cells (8) Added associations: 36
Bioinformatics	53 (17)	Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data. Bioinformatics is both an umbrella term for the body of biological studies that use computer programming as part of their methodology, as well as a reference to specific analysis "pipelines" that are repeatedly used, particularly in the fields of genetics and genomics. Common uses of bioinformatics include the identification of candidate genes and nucleotides (SNPs). Often, such identification is made with the aim of better understanding the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), or differences between populations. In a less formal way, bioinformatics also tries to understand the organisational principles within nucleic acid and protein sequences. (source: Wikipedia “bioinformatics”). Closely connected with: Synthetic biology (16), computational biology (8), biomimetics (7), supramolecular chemistry (5) Added associations: 36 Partly Overlaps with biosensors and complex systems (systems biology)

Plasmonics	50	Could be linked to photonics, because plasmonics refer to information transfer in nanoscale structures by means of surface plasmons. Similar things are being done in photonics. (Source: Wikipedia “surface plasmon”) However, as there is no light involved here but surface charges, and because there are so many single occurances of the keyword, we will keep this cluster separately. Closely connected with: - Added associations: -
Spintronics	37	Spintronics (a portmanteau meaning spin transport electronics, also known as spinelectronics or fluxtronics, is the study of the intrinsic spin of the electron and its associated magnetic moment, in addition to its fundamental electronic charge, in solid-state devices. Spintronics differs from the older magnetoelectronics, in that spins are manipulated by both magnetic and electrical fields. (source: Wikipedia “spintronics”) Closely connected with: - Added associations: - Same as plasmonics this is a stand-alone cluster, because it is unclear whether or not it shall be grouped into the nanotechnology-cluster.
Keywords not to be grouped in any of the above clusters because they are: -emerging research fields with no clear profile or disciplinary association - unclear what they mean - a prefix that was separated from the keyword it belongs to (like “real-time”)		-real-time (32) - trust (20) - sustainability (15) - tissue engineering (12) - social signal processing (8) - self-searching (8) - mobile devices (8) - mobile services (6)

The numbers assigned to the cluster are only indications for the strength of the cluster within the sample. In order to assess the selectiveness of the generated clusters, we checked how many proposals would in fact fall into the 19 clusters respectively how many proposals would use combinations of keywords from different clusters. The result of this cross-check was that we have indeed compiled very selective clusters: Only 174 proposals (4% of the whole sample) show overlaps at the keyword-level. This means in turn that 96% of all proposals are non-ambiguously assignable to one of the found 19 clusters (100%=4.035 proposals).

Results of the keyword analysis summarized in 19 clusters

Based on these results we described what is in the clusters in an observe-like way and fed these clusters into the Horizon Scanning database. We also checked the clusters with topics from the different FET Proactive calls to give an indication on whether these topics had also been identified by FET as future topics.

The 19 clusters are ranked according to their importance in the FET proposals´ sample. Table 5 shows the results of the clustering exercise.

Table 5: Results of the clustering exercise based on the keyword analysis of FET proposals

1. Advances in machine learning will boost the usability of computers in daily life Speech recognition, pattern recognition, data-driven predictions and decisions are fields in which computers could be of high usefulness in daily life. However, the results of machine learning research were more or less disappointing in the past. This may change fundamentally in the near future because new data sources become available and advances in machine learning, multi-agent systems and computational modeling can be combined in order to build working devices and software which helps people to cope with complex sourroundings. “Machine learning” and subfields were ranked no 1/19 in the keyword-search of the FET-Open/ FET-Proactive database.
2. Nanotechnology will finally deliver groundbreaking results Knowledge about nanoscale technologies has been steadily growing over the last years and some new applications as a result of heavily funded nanoresearch were developed. However, there was no spectacular breakthrough or broad diffusion of incremental advances sofar. But as soon as some basic research questions are answered, we will see groundbreaking advances in nanoelectronics, semiconductor physics, microfabrication, nanomedicine, nanophotonics, organic chemistry, molecular biology, and a range of other fields. “Nanotechnology” and subfields were ranked no 2/19 in the keyword-search of the FET-Open/ FET-Proactive database. (Nanotechnology was called up in FET_Proactive workprogrammes)

3 Quantum technology will move from basic research to applications Quantum technology today is in an early stage of its development and research is often characterized by theoretical reflections. However, the promises concerning future applications are manifold and are not limited to the quantum computer, although many expectations of the quantum community lie on this. Further expected applications range from secure communications, highly sensitive sensors to other breakthroughs in the context of data processing. For some researchers it is clear that the 21. Century will be the century of quantum technology (see topic no. 12 in the bottom-up clustering of FET open projects in D 1.1). “Quantum technology” and subfields were ranked no 3/19 in the keyword-search of the FET-Open/ FET-Proactive database. (Quantum technology was called up in FET_Proactive workprogrammes)
4. Complex Systems: Mathematical formalization will lead to a better understanding of our behaviour and surroundings. The study of complex systems represents a new approach to science that investigates how relationships between parts give rise to the collective behaviours of a system and how the system interacts with its environment. Since complex systems have many interconnected components, network theory is an important aspect of the study of complex systems. Fields that draw from these advances in this research area are swarm intelligence and swarm robotics, self-organising systems, adaptive systems and evolutionary computation. But also systems biology and affective computing make use of a better understanding of the relationships of elements in complex systems. “Complex Systems” and subfields were ranked no 4/19 in the keyword-search of the FET-Open/ FET-Proactive database. (Complexity was called up in FET_Proactive workprogrammes)
5. The boom of social media brings back the “social” into social and communication science. The social network perspective provides a set of methods for analyzing the structure of whole social entities as well as a variety of theories explaining the patterns observed in these structures. The methods are used to identify local and global patterns, locate influential entities, and examine network dynamics. Paradoxically, by using computer generated data resp. data from technically mediated communications, the social nature of all human communication and behavior is being emphasized. Future research will thus be much stronger oriented towards the social aspects than was the case in earlier times. This includes the analysis of social media, crowdsourcing, serious games, information retrieval, semantic web and others. “Social Networks” and subfields were ranked no 5/19 in the keyword-search of the FET-Open/ FET-Proactive database.

6. Computational Neuroscience holds the key for a better understanding of the human brain and cognition. Neuroscience has become an interdisciplinary science per se, combining knowledge and methods from chemistry, medicine, biology, including genetics, cognitive science, and linguistics as well as engineering, mathematics and computer science. Because novel neuorimaging technologies are generating masses of data, computer scientists and mathematicians need to develop new models and programmes to analyse this data. Either within experimental settings in vitro or in the context of medical research Computational Neuroscience is believed to have the potential to lead to a breakthroughs in the different neuroscience research fields. “Neuroscience” and subfields were ranked no 6/19 in the keyword-search of the FET-Open/ FET-Proactive database. (Computational Neuroscience was called up in FET_Proactive workprogrammes)
7. Research in Cloud Computing will lead to a world of fully converged infrastructures and shared services. Cloud computing is a kind of Internet-based computing, where shared resources and information are provided to computers and other devices on-demand. It is a model for enabling ubiquitous, on-demand access to computers and mobile devices. Research for cloud computing, storage solutions, ubiquitous and pervasive computing will speed up the current development where computing power and software is not necessarily located at the end-users device but available online and in realt-time for every device connected to the Internet. Advances in software engineering and computer science are necessary for fully converged infrastructures (Body Area Networks, Local Area Networks, fixed networks, backbone networks, etc.) and ubiquitously shared services. “Cloud Computing” and subfields were ranked no 7/19 in the keyword-search of the FET-Open/ FET-Proactive database.
8. New kinds of sensors (biosensors, lab-on-a-chip, smart textiles, etc.) and their smart connection will give us a new level of control over our surroundings. The reason why the Internet of Things still seems to be distant promise today is that there is not enough smart and connected data available. Thus, research needs to be encouraged in the area of sensing devices and technologies to analyse and exploit the data collected in the various fields. This research includes new kinds of biosensors, lab-on-a-chip technologies, wireless sensor networks, remote sensing, environmental monitoring, smart textiles and RFID. “Biosensor/sensing” and subfields were ranked no 8/19 in the keyword-search of the FET-Open/ FET-Proactive database.

9. Research in Human-Computer-Interaction will generate novel user interfaces and media applications which will revolutionize computer use in the future Human–Computer_Interaction (HCI) researches the design and use of computer technology, focusing particularly on the interfaces between people and computers. As a field of research, HCI is situated at the intersection of computer science, behavioral sciences, design, media studies, and others. Researchers believe that the future use of computers will be more intuitive, more ubiquitous, more haptic and multimodal. And it will intervene much stronger into the human perception of reality as it will be more immersive. In addition, advances in HCI will also advance our understanding of human-to-human interaction because of the need to formalize interactions. (Related to “robotics”) “Human-Computer-Interaction” and subfields were ranked no 9/19 in the keyword-search of the FET-Open/ FET-Proactive database. (HCI was called up in FET_Proactive workprogrammes)
10. Groundwork for Artificial Intelligence is still necessary despite of some useful applications Artificial Intelligence (AI) is the “intelligence” exhibited by machines or software. Although some have already termed the 21st century as the century of AI and some useful applications like “intelligent” Internet search machines or recommendation features have been developed, there are still more promises than real-life applications. One of the reasons is that the subfields of AI often fail to communicate with each other. For the time being, groundwork research will have to be done for example in the fields of ontologies, knowledge representation, computational linguistics and cognitive systems. “Artificial Intelligence” and subfields were ranked no 10/19 in the keyword-search of the FET-Open/ FET-Proactive database. (AI was called up in FET_Proactive workprogrammes)

11. Research in the field of photonic crystals may lead to a superior-performance optical computer. A photonic crystal is a periodic optical nanostructure that affects the motion of photons in much the same way that ionic lattices affect electrons in solids. Photonic crystals occur in nature in the form of structural coloration—and, in different forms (one-, two- or three-dimensional), promise to be useful in a range of applications. Photonic crystals can, in principle, find uses wherever light must be manipulated. Existing applications include thin-film optics with coatings for lenses. Two-dimensional photonic-crystal fibers are used in nonlinear devices and to guide exotic wavelengths. Three-dimensional crystals may one day be used in optical computers with superior performance. The related research field are silicon photonics, spectroscopy, optoelectronics, GaN, integrated optics, opto-mechanics, and photonic integrated circuits. “Photonic Crystal” and subfields were ranked no 11/19 in the keyword-search of the FET-Open/ FET-Proactive database.
12. Terahertz communication enables a new range of wireless applications in the future In May 2012, a team of researchers from the Tokyo Institute of Technology set a new record for wireless data transmission by using Terahertz-rays and proposed they be used as bandwidth for data transmission in the future. The new devices will use a so called resonant tunneling diode (RTD) in which the voltage is decreased as the current increased, causing the diode produce waves in the terahertz band which enable data transfer rates of 3 Gigabits per second. The demonstration was twenty times faster than the current Wi-Fi standard. Research still has to be done, especially in the fields of adaptive antennas, frequency comb to unleash the potential of the invention. “Terahertz” and subfields were ranked no 12/19 in the keyword-search of the FET-Open/ FET-Proactive database.
13. Interdisciplinary research to build context-aware robots Robotics is the branch of electronic and electrical engineering and mechanical engineering and computer science that deals with the design, construction and operation of robots, as well as computer systems for their control, sensory feedback, and information processing. Wheras robots are already in use in production environments and other contexts where simple and fixed tasks need to be done, autonomous, rule-based and context-sensitive “behavior” is currently the dominant research challenge in robotics. Thus, more research will be needed in the future in the areas of context awareness, embodied cognition, autonomous systems, and human-robot-interaction. “Robotics” and subfields were ranked no 13/19 in the keyword-search of the FET-Open/ FET-Proactive database. (Robotics was called up in FET_Proactive workprogrammes)

14. CMOS technology will stay on the R&D agenda and will continue to boost chip performance and bring about new applications Complementary metal–oxide–semiconductor (CMOS) is a technology for constructing integrated circuits. CMOS technology is used in microprocessors, microcontrollers, static RAM, and other digital logic circuits. CMOS technology is also used for several analog circuits such as image sensors (CMOS sensor), data converters, and highly integrated transceivers for many types of communication. Research will continue in this area and new performance output and new applications will come out of this R&D activities in the future. "CMOS" and subfields were ranked no 143/19 in the keyword-search of the FET-Open/ FET-Proactive database. (CMOS and nanoelectronics were called up in FET_Proactive workprogrammes)
15. In the research field of Microfluidics, interdisciplinary work will yield advanced Lab-on-a-Chip-technologies and other new applications Microfluidics is a multidisciplinary field intersecting engineering, physics, chemistry, bio-chemistry, nanotechnology, and biotechnology, with practical applications to the design of systems in which low volumes of fluids are processed for high-throughput screening. Typically, fluids that are constrained to a sub-milimeter scale are moved, mixed or sepa-rated. Thus, processes which are normally carried out in a lab can be miniaturized on a single chip (lab-on-a-chip). Other current applications of microfluidics include inkjet printheads, DNA-chips, micro propulsion and micro-thermal technologies. With the intro-duction of active micro components like micropums or microvalves, a whole new range of applications will become possible in the future. "Microfuilics" and subfields were ranked no 15/19 in the keyword-search of the FET-Open/ FET-Proactive database. (Microfluidics was called up in FET_Proactive workprogrammes)

16. Energy Harvesting will be the solution for powering small and mobile devices in the future. Energy harvesting is the process by which energy is derived from external sources (e.g. solar power, thermal energy, wind energy, salinity gradients, and kinetic energy), captured, and stored for Small, wireless autonomous devices, like those used in wearable electronics and wireless sensor networks will more and more be powered by energy from kinetic energy, salinity gradients, solar power or thermal energy instead of batteries. The process of making these external energy sources usable for small devices is called “Energy harvesting”. Once new and efficient ways of harvesting energy from the environment have been found, more and more mobile devices will be powered by sourrounding renewable energy sources. Also, once successful, more and more low-energy electronics will be developed and deployed in a variety of fields. “Energy harvesting” and subfields were ranked no 16/19 in the keyword-search of the FET-Open/ FET-Proactive database. (Minimizing energy consumption for computers was called up in FET_Proactive workprogrammes)
17. Bioinformatics: New software tools will help solving the big issues in biology Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data, especially in genetics and genomics. Common uses of bioinformatics include the identification of candidate genes and nucleotides which aims at a better understanding of the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), or differences between populations. Bioinformatics contributes to advances in synthetic biology biomimetics, supramolecular chemistry and other subfields. As such, researchers expect bioinformatics to provide major tools and methods to solve basic questions of biology and genetics. “Bioinformatics” and subfields were ranked no 17/19 in the keyword-search of the FET-Open/ FET-Proactive database. (Bioinformatics was called up in FET_Proactive workprogrammes, namely as research in the field of bio-inspired intelligent information systems)

18. Plasmonics: From basic resarch to breakthroughs in high-performance computing and nano devices. Plasmonics deal with information transfer in nanoscale structures, similar to photonics, by means of surface plasmons. In contrast to photonics, there is no light involved but surface charges are being analysed in this research field. Being in an early stage of development, plasmonics may contribute to new high-performance chips and nano devices since surface plasmon-based circuits can overcome the size limitations of photonic circuits. “Plasmonics” was ranked no 18/19 in the keyword-search of the FET-Open/ FET-Proactive database.
19. Spintronics: **New principles for new, ultra-high capacity storage devices**. Sprintronics is a research field of nano electronics and deals with the study of the intrinsic spin of the electron and its associated magnetic moment. Spintronics differ from older magnetoelectronics, in that spins are also manipulated by magnetic fields and not by electrical fields alone. One promise of this research field, which is at the same time basic research and applied research, is that with the help of the special principles, storage devices with much higher capacities may be developed in the future. “Spintronics” was ranked no 19/19 in the keyword-search of the FET-Open/ FET-Proactive database.

3.4 Web-Mining

One element of the Observe trend analysis was to use web content as a data source to recognize emerging themes. The web mining part builds on the analysis of social media data and web content. So far, web analysis is not frequently applied in foresight. In the scientific literature, some applications use web mining to screen company web sites (innovation indicators). (Cachia et al. 2007) evaluate the potential of online social networks for foresight and trend recognition. They conclude that social networks indicate changes and trends in sentiment and social behavior and among other aspects serve for debates of different future developments. (Pang 2010a) develops an approach to scanning of web 2.0 contents produced by futurists on different web channels. This could deliver a very precise summary of what is discussed online and what attracts futurist’s attention (Pang 2010b). (Amanatidou et al. 2012) use Twitter and other publicly available web sources and conclude that it is a great channel for real time information but needs good filtering (precise tag selection and reliable senders). To process the volume of information published, web and text mining are applied. This can help in identifying trends and newest developments in an objective manner and recognize developments that are overlooked otherwise.

Implementation

The social media platform Twitter was used for data gathering. Data was aggregated from trend blogs and the trend and technology section of newspapers. In total 27 channels were selected that are characterized by a high degree of future orientation. On 24th September 2015 for each channel the last 3500 tweets[15] were retrieved. From these, retweets and replies were excluded so that only original tweets were analysed. This finally led to a sample of 70250 Tweets. As some of the channels tweet only rarely, this set also included some tweets from previous years but the large majority (53004) were from 2015. The following table gives an overview on the number of tweets per channel.

Table 2: Twitter channels monitored in the webmining

No.	Channel	Number of original Tweets analysed in the OBSERVE webmining
1	BBC SciTech	3200
2	NYT Technology	3200
3	Gizmag	3196
4	Guardian Science	3110
5	TechCrunch	3103
6	Guardian Tech	3064
7	Technology Review	3005
8	The Economist	2988
9	WIRED Science	2945
10	HuffPost Tech	2920
11	Forbes Tech News	2918
12	The Verge	2854
13	New Scientist	2796
14	Mashable Tech	2763
15	WIRED	2756

15 Technical restriction of Twitter Platform

16	GuardianScienceBlogs	2738
17	BBC Technology	2614
18	NYT Science	2553
19	BBC Future	2549
20	TED News	2433
21	XPRIZE	2382
22	MIT Tech Review	2366
23	Science Magazine	2287
24	The Royal Society	1966
25	Nesta	1718
26	Deep Stuff	1698
27	science tracker	128

Analysis of hashtags

Around 11.000 of the collected tweets contained hashtags aligning them to certain thematic lines of twitter discussion. To analyze the hashtags contained in the tweets, first a frequency analysis was conducted. The analysis however revealed no interesting insights in terms of content as most hashtags were either very general (#science), referring to the type of content (#free) or very specific events (#solareclipse, #rethinking-parks) or products (#apple). Several of the most frequent hashtags however were **conference tags** such as e.g. #emtechmit. The hashtaged sites listed below were manually reviewed for interesting emerging topics.

Table 3: Conferences emerging from the webmining

Hashtag	**Conference**
#summerscience	The Summer Science Exhibition is an annual display of the most exciting cutting-edge science and technology in the UK. This week-long festival features 22 exhibits from the forefront of innovation. 30 June - 5 July 2015, London
#emtechmit	EmTechMIT Conference of the MIT Technology Review Magazine featuring 10 breakthrough technologies and innovators under 35. November 2–4, 2015, MIT Media Lab, Cambridge, MA

#rsbioinspire	Conference of the Royal Society: Bioinspiration of New Technologies. London, May 2015

Text mining the tweets

In a next step the actual text of the tweets was further analysed. A first cluster analysis[16] revealed the main topics discussed in the twitter corpus. Of these topics the ones considered most relevant for the context of OBSERVE (Energy, Solar Energy, Sea/Ocean/Water, Brain/Neuro, Robot, Light) were then extracted and further analysed manually for emerging topics. As an example the analysis of the subset of tweets on brain discourse revealed the following aspects presented in Box 1.

Box 1: Emerging topics revealed by the analysis of the brain related subset of tweets

- Understanding the way the brain works in particular the memory but also spatial mapping, timing, vision, decision making, emotional experience assignment, social prediction, hearing, tinnitus, pattern recognition and aging.
- New imaging techniques for observing the brain at work such as deep brain imaging
- Directly connecting brains (synthetic telepathy)
- Brain mimicry for neuromorph computing especially memristors and optical fibres for brain circuits, ethical questions on the ethical status of simulated brains in the far future
- Brain implants: Soft devices to deliver drugs in the brain, syringe-injectable electronics, interfaces for controlling (bionic) prostheses through thoughts
- Non-invasive influencing of brain cells: Through sound waves, through magnetic stimulation (using nanoparticles)
- Brain training especially for people with schizophrenia
- Creation of an artificial brain from stemcells
- Evolution of the brain
- Detecting brain injuries through blood tests via biomarkers

---

16 The analysis was based on counting the frequency of terms and comparing it to the overall structure of the corpus. We thank Prof. Octavian Popescu for his support in this task.

Some of these aspects were integrated into already existing clusters from the manual scouting (e.g. brain implants) other became new entries (e.g. synthetic telepathy). The same proceeding was adopted for the other domains.

In total the webmining led to 15 new entries of which seven became clusters. Many others were integrated into the existing clusters from the manual scouting. In particular the list of unconventional new energy sources was enlarged by the webmining analysis.

3.5 Publication Analysis

Concept

Scientific publications are an essential output of scientific work. Their mathematical and statistical analysis reveals insights on developments, dynamics and new research directions within different fields. Accordingly, bibliometric analysis has long been used to map dynamics of change in specific domains. Until now however there are only few attempts to use bibliometric analysis for identifying newly emerging S&T topics without a preexisting framework (Small et al. 2014). In their review of definitions and techniques, (Cozzens et al. 2010) report that most studies of emerging technologies are retrospective analyses of pre-determined areas rather than methodological studies designed to identify emerging technologies whereas identification of emerging topics remains a challenge. Other studies screen the full S&T landscape for those papers with the highest number of citations in order to identify “hot topics”. Such an approach however is not suitable in this case as it reveals dominant topics in the scientific debate rather than novel ideas that differ from the existing landscape and may thus not yet receive a lot of attention.

In our approach we will build on a very recent strand of theory using cluster assessment in sliding time windows to identify newly emerging topics such as Small et al (2014), (Michels, Rettinger 2014), (Huang, Chang 2014). In this approach the publications are clustered within a certain time-span such as e.g. a three year periods that are moved upward step by step (“sliding windows”). New clusters are identified by using a similarity indicator between time periods. The clusters with the smallest maximum similarity are chosen as emerging topic candidates. For forming the clusters there are different approaches. Some authors such as (Boyack, Klavans 2010) and (Small et al. 2014) use co-citation analysis i.e. grouping papers that cite each other into clusters. Other research teams use bibliographic coupling i.e. clustering papers according to their references so that papers that share the same references are grouped together Huang and Chang (2014), Michels and Rettinger (2014). In OBSERVE we will use bibliographic coupling for two reasons. Firstly, citation analysis involves a certain time delay as new papers are only cited a few years after their first publication. This is critical for the case of OBSERVE, where we target very recent developments. Secondly, as the aim of the FET call is also to identify new potential collaboration opportunities it is of particular interest to identify groups of papers that share a knowledge base without necessarily knowing of and citing each other. After having clustered the full set of publications within each year from 2004-2014, recent new clusters will be identified by using a similarity indicator. The clusters with the smallest maximum similarity are chosen as emerging topic candidates, since these are the topics that are most likely to be new. We have tested the feasibility of this approach as described in detail by (Michels and Rettinger 2014) and were able to show that it yields relevant results.

The data for the bibliometric analysis is retrieved from the Web of Science™ Database. The Web of Science™ Core Collection provides access to the world's leading citation databases. Authoritative, multidisciplinary content covers over 12.000 of the highest impact journals worldwide—including open access journals—and over 160,000 conference proceedings. Focus on the essential data from across more than 250 disciplines and find information in scientific areas such as agricultural, biological, and environmental sciences, engineering, technology, applied science, medical and life sciences, and physical and chemical sciences.

Implementation

The bibliographic coupling algorithm was programmed using an overlap of 30% of references as criterion for forming clusters and started to analyse the dataset of publications from the Web of Science™ Database[17] in the sliding window of four years as described above. Unfortunately, it turned out that the calculations took a very long time and could not reasonably be finalised within the OBSERVE scouting period. In order not to delay further the project it was decided to feed in these results at a later point in time.

In the meantime for the scouting phase we integrated the results of the Research Fronts 2014 published by Thomson Reuters in collaboration with the Chinese Academy of Sciences in December 2014 (The National Science Library Chinese Academy of Sciences et al. 2014). The publication provides a total of 144 research fronts in 10 broad domains that were identified using co-citation analysis as described above. A research front consists of a set of a highly cited papers and the respective group of citing papers. For each domain 10 "hot research fronts" are provided i.e. the most re-

---

[17] Update 2015

cent and most cited front within the area. The 44 emerging fronts are the most recent ones across all areas. As the focus of this approach is very much on the attention the papers receive in the scientific community it is not perfect for identifying novel and interdisciplinary topics that may well receive little attention at first. Also, the time lag is considerably as at least a year is needed before a paper can accumulate many citations. Nevertheless, the emerging research fronts very nicely complement the insights from the OBSERVE manual scouting and Webmining which were focussing on more recent and unconventional debates.

From the 144 research fronts we selected the ones from domains relevant for FET funding i.e. excluding the ones in basic research areas like physics, biology, astronomy and medical sciences. Also we did not include any fronts with core papers older than 2011. In total 19 research fronts were captured within the OBSERVE radar including 9 emerging research and 10 hot research fronts.

4 Lessons learned/Recommendations

4.1 Scouting Framework

The main challenge of the scouting process was the development of the screening framework. Naturally, with more experience along the process the categories were also evolving which meant that entries needed to be reassessed several times along the process. Another challenge was the forming of clusters. Whereas in the beginning the aim of the clustering was merely to combine entries pointing out the same phenomenon in order to reduce the number of entries, some of the clusters became broader along the way e.g. mobility futures containing very different developments in the area of mobility. In some cases the challenging aspect of the original entry may have been lost in this process. At the same time, sticking to all original entries would have exploded the number of entries some of them on a very micro level which would have greatly complicated the synthesis process. Related to this challenge is the general problem of defining an adequate level of granularity. As it is now the radar contains very small things (e.g. development of a foldable microscope) to very large issues like “future of civilisation”. This diversity is not necessarily a problem for the use of the radar in OBSERVE as items with different levels can later be combined into hotspots.

4.2 Manual Scouting

Key for the quality of the manual scouting is the selection of sources, the sensitivity of the scouts and a good shared understanding of the type of finding that is targeted. Both aspects evolved considerably within the scouting phase. We would therefore recommend for similar exercises in the future foreseeing a pre-scouting phase of at least three months for fine-tuning the set of sources and screening criteria.

The analysis of science fiction literature was less fruitful than expected as several of the items mentioned there did not go much beyond current technology roadmaps. Also the analysis of the Kick-Starter platform generated mostly information about topics of public interest but less about newly emerging developments.

One problematic issue is the language restriction. Even though we have made an effort to select sources with a high degree of international orientation the restriction toe English (and German) language sources may have been a limiting factor.

4.3 Keyword analysis of FET proposals

The analysis of research proposals in the FET Open and FET Proactive scheme was based on the assumption that these proposals can be used as indicators for new and evolving topics in research and technology. Looking at the results (the 17 findings ranging from to machine learning to nanotechnology and bioinformatics), there is a plausibility for this assumption as the finding are indeed interesting and convincing. However, since there is a certain congruence of the clusters found with the topics supported within the FET Open as well as the FET Proactive scheme, the results of this method might be disappointing for someone expecting big surprises. In fact, the method of clustering used here systematically levels out very special approaches or unseen topics which are generally of interest as well. On the other hand, what is won by using this method is a systematic approach which hints to different topics in research and technology development which are currently considered as new and emerging, and which require interdisciplinary and collaborative approaches according to the assessment of a high number of researchers.

The other issue is the question of automatisation. The process of clustering the 4.000 proposals was designed as a semi-automated process which consisted of an automated counting of keyword frequencies and a manual clustering of topics. In fact, in the beginning, we expected to get to interesting and valid results with a fully automated process of counting keywords from the proposals only. This proved to be naïve to a certain extent: As described at the beginning of chapter 3.3 we started with the assumption that counting the words “new”, “first”, “unconventional”, and “innovative” in the proposals´ abstracts would identify topics which would be exactly that. A closer look at the results revealed that this might not be the case because the method takes proposers´ claims at face value and does not critically examine whether or not these claims actually hold in real life. Finally, we decided for a different method which also counted keyword frequencies but which required as a second step the manual examination of the results, e.g. a manual clustering (see section “Step 3: analysis of freely given keywords in a two-step clustering exercise”).

Before deciding on this specific clustering method we have tried several other ways – all of which were of higher automatisation. For example, after the first round of clustering we checked for the exclusiveness of the clusters and found very good values. We concluded that proposals belonging to more than one topical cluster must be research requiring interdisciplinary approaches per se. Also, the more overlaps we would find in a single proposal, the newer and more unconventional its approach seemed to be. When looking closer into the topics of these specific projects we found in fact very special set-ups. The topics followed in these projects were highly specialized and very singular. Thus, the attempt to describe these topics as findings for the observe database would have resulted in describing a set of highly specialized research from which we would not be able to say that they form a cluster at all. Therefore we decided to follow a different approach (as described in the Step 3-section). One major learning in this process was that automation of the analysis only carries to a certain level. From this level, expert knowledge and manual assessments become necessary in order to achieve plausible and reliable results.

4.4 Webmining

As explained above it emerged that analysis of hashtags from science and technology blogs was not useful for this purpose as the hashtags usually do not refer to the science and technology content. In contrary the analysis of the full text of the tweets revealed useful insights on the most important topics of the debate. Nevertheless, in order to derive the final contributions for the OBSERVE radar substantial manual analysis was required. In general it should be noted that a quantitative assessment of the tweets requires substantial upfront data processing as each channel has different “tweeting policies” so a high number of appearance of a certain term does not necessarily signify a high relevance. Also this method (even more than the manual scouting) is very sensitive to the general thrust of the public debate so e.g. events like the solar eclipse and the NASA Mars exploration sparked many space related tweets even in science and technology blogs without this being an indication of actual S&T breakthroughs.

4.5 Publication Analysis

As it is obvious from the above account we severely underestimated the calculation intensity of the biographical coupling account applied to the whole range of research fields. We will report on the solutions developed at a later stage in the project.

5 References

Amanatidou, E.; Butter, M.; Carabias, V.; Könnölä, T.; Leis, M.; Saritas, O.; Schaper-Rinkel, P.; van Rij, V. (2012): On concepts and methods in horizon scanning: Lessons from initiating policy dialogues on emerging issues. In: Science and Public Policy, 39 (2), pp. 208-221.

Ansoff, H.I. (1975): Managing strategic surprise by response to weak signals. In: California Management Review, XVIII, pp. 21-33.

Bijker, W.E.; Law, J. (1997): Shaping technology / building society. Studies in socio-technical change, Inside technology, Cambridge/Mass.: MIT Press.

Boyack, K.W.; Klavans, R. (2010): Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? In: Journal of the American Society for Information Science and Technology, 61 (12), pp. 2389-2404.

Cachia, R.; Compañó, R.; Da Costa, O. (2007): Grasping the potential of online social networks for foresight. In: Technological Forecasting and Social Change, 74 (8), pp. 1179-1203.

Cozzens, S.; Gatchair, S.; Kang, J.; Kim, K.S.; Lee, H.J.; Ord+¦+¦ez, G.; Porter, A. (2010): Emerging technologies: quantitative identification and measurement. In: Technology Analysis & Strategic Management, 22 (3), pp. 361-376.

Cuhls, K.; van der Giessen, A.; Toivanen, H.; Erdmann, L.; Warnke, P.; Toivanen, M.; Seiffert, L. (2015): Models of Horizon Scanning. How to integrate Horizon Scanning into European Research and Innovation Policies.

Day, G.S.; Schoemaker, P.J. (2004): Peripheral Vision: Sensing and Acting on Weak Signals. In: Long Range Planning, 37 (2), pp. 111-121.

Ecken, P.; Gnatzy, T.; von der Gracht, H.A. (2011): Desirability bias in foresight: Consequences for decision quality based on Delphi results. In: Technological Forecasting and Social Change, 78 (9), pp. 1654-1670.

Huang, M.H.; Chang, C.P. (2014): Detecting research fronts in OLED field using bibliographic coupling with sliding window. In: Scientometrics, 98 (3), pp. 1721-1744.

Ilmola, L.; Kuusi, O. (2006): Filters of weak signals hinder foresight: Monitoring weak signals efficiently in corporate decision-making. In: Futures, 38 (8), pp. 908-924.

Könnölä, T.; Brummer, V.; Salo, A. (2007): Diversity in foresight: Insights from the fostering of innovation ideas. In: Technological Forecasting and Social Change, 74 (5), pp. 608-626.

Markley, O. (2011): A new methodology for anticipating STEEP surprises. In: Technological Forecasting and Social Change, 78 (6), pp. 1079-1097.

Michels, C.; Rettinger, A. (2014): Emerging Topics in Science - Subproject in the Kompetenzzentrum Bibliometrie, Fraunhofer ISI (ed.), Discussion Papers Innovation Systems and Policy Analysis No. 42, Karlsruhe.

Neugarten, M.L. (2006): Foresight: Are we looking in the right direction? In: Futures, 38 (8), pp. 894-907.

Pang, A.S.-K. (2010a): Futures 2.0: rethinking the discipline. In: Foresight, 12 (1), pp. 5-20.

Pang, A.S.-K. (2010b): Social scanning: Improving futures through Web 2.0; or, finally a use for twitter. In: Futures, 42 (10), pp. 1222-1230.

Rossel, P. (2012): Early detection, warnings, weak signals and seeds of change: A turbulent domain of futures studies. In: Futures, 44 (3), pp. 229-239.

Schoemaker, P.J. (2003): Organizational Renewal: Overcoming Mental Blindspots. The Many Facets of Leadership, New Jersey: M.Goldsmith.

Small, H.; Boyack, K.W.; Klavans, R. (2014): Identifying emerging topics in science and technology. In: Research Policy, 43 (8), pp. 1450-1467.

The National Science Library Chinese Academy of Sciences; Thomson Reuters IP & Science; The Joint Research Center of Emerging Technology Anaylsis (2014): Research Fronts 2014: 100 Top ranked specialities in the sciences and social sciences.

von Hippel, E. (2006): Democratizing Innovation: The evolving phenomenon of User Innovation. In: Kahin, B.; Foray, D. (eds.): Advancing Knowledge and the Knowledge Economy. pp. 237-255.

Warnke, P.; Schirrmeister, E. (2016): Small seeds for grand challenges. Exploring disregarded seeds of change in a foresight process for RTI policy. In: Futures, 77, pp. 1-10.