Eur. Phys. J. Special Topics **214**, 347–360 (2012)

DOI: 10.1140/epjst/e2012-01698-7

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

Exploratory of society

L.-E. Cederman[1], R. Conte[2], D. Helbing[3], A. Nowak[4], F. Schweitzer[5], and A. Vespignani[6]

[1] International Conflict Research at ETH Zurich, Zurich, Switzerland
[2] National Research Council Institute of Cognitive Science and Technology, Roma, Italy
[3] Risk Center, ETH Zurich, Zurich, Switzerland
[4] Warsaw School of Social Sciences and Humanities in Warsaw, Poland
[5] Chair of Systems Design, ETH Zurich, Zurich, Switzerland
[6] Northeastern University, Boston, MA, USA & ISI foundation, Turin, Italy

Received in final form 9 October 2012
Published online 5 December 2012

Abstract. A huge flow of quantitative social, demographic and behavioral data is becoming available that traces the activities and interactions of individuals, social patterns, transportation infrastructures and travel fluxes. This has caused, together with innovative computational techniques and methods for modeling social actions in hybrid (natural and artificial) societies, a qualitative change in the ways we model socio-technical systems. For the first time, society can be studied in a comprehensive fashion that addresses social and behavioral complexity. In other words we are in the position to envision the development of large data and computational cyber infrastructure defining an exploratory of society that provides quantitative anticipatory, explanatory and scenario analysis capabilities ranging from emerging infectious disease to conflict and crime surges. The goal of the exploratory of society is to provide the basic infrastructure embedding the framework of tools and knowledge needed for the design of forecast/anticipatory/crisis management approaches to socio technical systems, supporting future decision making procedures by accelerating the scientific cycle that goes from data generation to predictions.

1 Vision

Human societies encompass numerous activities in economics, technology, infrastructure, society, and environment, which cannot be separated from another. Hence, an Exploratory of Society has to develop interfaces with exploratories in all those areas, to consider data transfer and feedback between them. Specifically, the exploratory of Society will initially concentrate on four dimensions of society:

– Crime and Conflicts.
– Health and Contagion.
– Social Resilience.
– Social Well-being.

For each exploratory, we need to (i) define what kind of data should be made available in the four areas mentioned, and (ii) how such data is integrated and aggregated, to allow for useful insight. This involves both an engineering component regarding data gathering, data aggregation, efficient data processing, and a theory-driven operational criteria and the relative measures that only provide such insight and the models that allow forecasting, i.e., we have to provide tools and methods, but also questions and ideas of what should be explored in a society. FuturICT is proposing a novel architecture for the exploratories of society that defines at the same time a laboratory generating knowledge and defining experiments and a crisis and anticipation infrastructure [1]. This architecture can be realized only through a shift of paradigm and methodology that consist in:

- A community approach to the problem (truly interdisciplinary).
- Methods dealing with complex realities (Agent Based Models, Game theoretical models, Complex networks etc.).
- Integration of novel experimental behavioral science, cognitive science and neuroscience.
- Computational thinking: integration of expertise, techniques, and methodologies for computational and data-intensive science.
- Innovative use of Information and Communication Technologies (ICT) to gain knowledge on human, social and economical systems.
- Novel integration of social science (computational social science, predictive models, quantitative models).

The FuturICT exploratory of society build on previous efforts and successes showing that it is possible to produce non-incremental advances in the development of:

- Novel tools and computational models that can support policy makers.
- Tools for real-time detection of global trends and assessment of systemic risk.
- Model-driven contingency planning and scenario analysis.
- Computational coordination of required resources, logistic support and other associated processes in the context of crisis prevention and mitigation.

1.1 Goals and challenges

The goal of FuturICT is to enable purposeful action by societies in service of goal achievement and realization of its values, avoiding crisis and facilitating sustained growth. FuturICTs new concepts for the measurement, simulation and interactive exploration of the impact of human decisions and actions will promote collective awareness. Simulations of societal processes converging on the Global Earth simulator [2] will provide integrated anticipatory functions to humanity enabling to base decisions on considerations of their consequences. The crisis avoiding mechanisms will perform regulatory function in service of societies where some elements will performed by the computer system and some will just warn policy makers about possible threats and provide information needed for decision.

In the above context the exploratory of society defines the basic infrastructure embedding the framework of tools and knowledge needed for the design of forecast/anticipatory/crisis management approaches to a variety of critical domains that interfere positively and negatively within socio-technical systems, such as conflict, crime, as well as social and biological contagion. Approaches will include participatory policy modeling platforms, allowing policies and governance strategies to be designed by means of a coordinated collaborative effort from different stakeholders.

The vision of a working exploratory of society infrastructure requires an integrated approach to data acquisition and analysis, modeling, computational infrastructures that builds on the results of previous projects and systematically redesigns the collaborative interaction among the stakeholders of the various activities. The modeling frameworks need to be enlarged in order to deal with the complex features of present techno-social systems. This implies the use and development of techniques dealing with complex systems features as well as developing new theories, models and mathematical tools able to attain quantitative analytical power in large scale techno-social systems. The challenge is therefore to generate novel knowledge and theoretical understanding by processing previous acquired data. On the other hand theory-driven questions will trigger further data acquisition. Furthermore, as in weather forecasting, a large scale social exploratory infrastructure needs to rely on sophisticated computational tools to integrate present data and huge libraries of previous patterns into realistic modeling framework. Data gathering has to be informed on the modeling needs as models have to be refined according to the accessible data. The time scale of data acquisition and sharing should be considered as a key concept as it is in weather predictions; nobody would think to analyze weather on the basis of the meteorological situation of two months ago. Data and algorithm sharing with the scientific community at large, in real time, is an essential requirement. Finally the development of appropriate user interface that would define participatory policy modeling platforms, allowing policies and governance strategies to be designed by means of a coordinated collaborative effort from different stakeholders is necessary to make exploratories of society concrete and useful infrastructures. This calls for the development of novel ICT technologies.

1.2 State of the art

In the last years we have witnessed tremendous scientific and technological progress that has caused a qualitative change in the ways we model socio-technical systems [18,34]. Visualization and analysis tools able to cope with multiple levels of representation are being developed along with computer simulations that provide experiments not feasible in real social systems [4,6,20,24,31]. Although these computational experiments have to be validated in real world settings, they are the best tool to provide scenario analysis if based on massive high quality data sets. For this reason, a quest for innovative technologies comes from the real-time data gathering point of view. Innovative technologies can improve the traditional questionnaire-based systems, providing faster and better localized detection capabilities and resulting in a broad practical impact [10,16,24,27,32]. For the first time we are in the position to envision the development of large data and computational infrastructure defining exploratories of society, able to provide quantitative anticipatory and scenario analysis capabilities in phenomena ranging from emerging infectious disease spreading and urban mobility to conflict analysis and crime surges [3,5,14,19–21,29].

Crisis rooms around the world are already taking advantage of some of those novel capabilities [15,35]. However, it has become clear that any new and significant scientific progress to gain knowledge on human, social and economical systems requires a novel interdisciplinary integration of computational and data-intensive science with mathematical/computational modeling and an innovative use of Information and Communication Technologies (ICT) [18,25,34]. Many scientific projects especially in the FET program have pushed the envelope of scientific research in this area. However most of them take into consideration only one aspect of the problem without creating the appropriate research integration among data-modeling-monitoring efforts. Indeed, the vision of a working exploratory of society infrastructure does require an integrated approach to data and modeling that builds on the results of previous projects. FuturICT is therefore proposing a novel architecture that goes beyond the state of

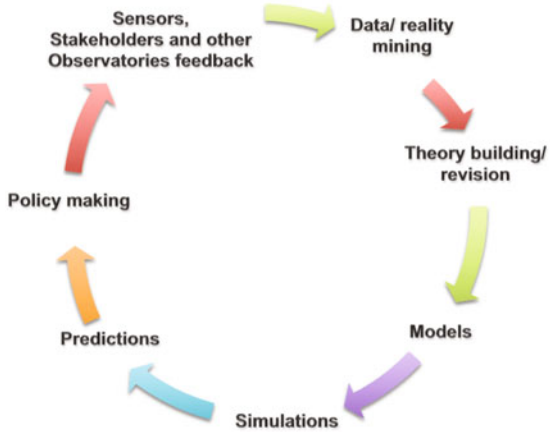

Fig. 1. Exploratory of society workflow.

the art of current crisis and situation rooms. In particular the exploratory of society is at the same time a laboratory generating knowledge and defining experiments and a crisis and anticipation infrastructure.

2 Proposed research strategy

The ambitious goal of an exploratory of society requires a multidisciplinary and interdisciplinary research approach that provides appropriate tools and knowledge. While we start by identifying a limited set of specific sub-areas such as global health, crime, conflict resolution etc., the final aim is an interconnected infrastructure where the various components are connected with additional exploratories developed within the FuturICT endeavor. The final goal of the exploratory of society is to provide useful technological tools and support for future decision making procedures by accelerating the workflow, shown in Fig. 1, that goes from data generation to predictions in a continuous feedback loop. This translates in a definite structure for the exploratory of society, reflected at the component level of the domain exploratories, where each step of the workflow can be mapped into a specific technical component that has to be developed, integrated and harmonized with the entire infrastructure as shown in Fig. 2. Here we provide a list of the main challenges for each of the component illustrated in Fig. 2.

Data acquisition and measurements

- Assimilating and integrating the ever-increasing wealth of datasets is needed to support the modeling approach and to extract knowledge and pattern from multiple data source.
- Unified and integrated approach for the management of these resources, with the design and implementation of standard sharing and management practices.
- For each specific exploratory, it is crucial to devise real time data gathering, integration and assimilation approaches (social adaptation, social distancing, evacuations, traffic etc.). We need to overcome the limitation of the state of the art surveillance/monitoring systems by proposing innovative ICT approaches.

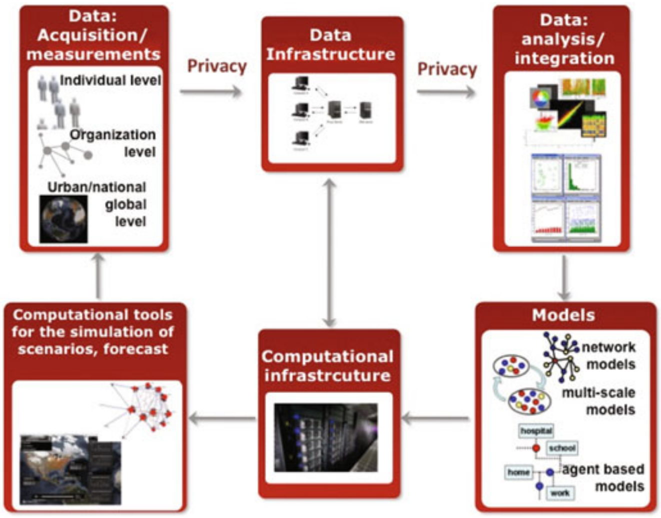

Fig. 2. Illustration of the technical components and work areas of the exploratories of society.

Privacy and ethical considerations

- The data of most exploratories will present huge privacy and ethical constraints the implementation of a full privacy preserving approach must be carefully implemented from the data source to the modeling and results analysis.
- Ethical considerations are not restricted to privacy matters. exploratories' activities must be inspired to much the same precepts that govern current ICT systems, in particular social responsibility. Both real-world and simulation-based experimental work should support and inform the dissemination activity.

Data infrastructure component

The data infrastructure is by definition delocalized according to different institutions, labs and collection points. It is however crucial (in data intensive application) to provide geographically localized collection points:

- They will support the sharing and management of datasets and resources as well as their rating, annotation, and selection through appropriate participatory platforms.
- They will be physically located in the same place of the computational infrastructures running data driven models (reduction of computational overhead, data manageability, etc.).

Data analysis/integration and discovery

- The identification of general principles and laws, extraction of knowledge, data reduction etc. that characterize complexity and capture the essence of the system.
- Comparison and empirical evaluation of sets of competing scenario analyses based on alternative combinations of models of fundamental social mechanisms driving societal macro dynamics.
- Results have to be integrated and feed the modeling component.

Modeling component

Processes and data studied need the development of foundational models and theories capable to provide a conceptual and understanding framework that addresses the complexity inherent to the biological, social and technological aspects of the problem at hand.

- Complexity and multi-scale nature of the techno-social systems.
- Heterogeneities in population aggregation and spatial structuring.
- Complexity of social dynamics and systems.
- Effect of heterogeneities on many magnitudes and spatial scales.
- Adaption and extension of approaches for large scale systems (10^4 to 10^8 degrees of freedom).
- New model structures and parameter estimation schemes.

Computational infrastructure

Data-driven models are integrated with massive amount of data. For many applications they need to run in real time for calibration, projections and merging. Models portfolios are run contextually to produce ensemble forecasting, model output statistics etc. The computational infrastructure also produces visualization outputs on and results processing all in real time to be streamed to crisis situation rooms around the world. Please note that not all the exploratories need this real time computational power. Other solutions concerns cloud computing and grid access points.

Data driven computational tools

The final aim is to provide flexible and user-friendly tools for the simulation of case studies, understanding of observed patterns, study of the effectiveness and results of different intervention policies, analysis of risk through model scenarios and forecast for the system of interest. This implies the implementation of data driven approaches endorsed with quantitative and predictive powers. Computational tools have to provide appropriate user interface for the control of model parameterization and the exploitation of the results from non expert (stakeholders, policy makers etc.). Multiple modeling comparison, model output statistics and ensemble analysis shall be implemented. Cross-methodological tools, including laboratory experimental platforms, should be also made available.

2.1 Success stories and demonstrators areas

True progress in knowledge production and understanding socio-technical systems require a systematic redesign of the research approach in terms of integration, multidisciplinary effort and collaborative interactions among the stakeholders of the various activities, building on the results, experience and successes reached in past and ongoing projects. For this reason we can consider some demonstrator areas as the building blocks of the Exploratory of society.

Health and contagion exploratory

All around the world 15 million deaths per year are estimated to be directly related to infectious diseases. The emergence of new diseases such as the severe acute respiratory syndrome (SARS), and the threat of new influenza pandemic are just two examples of the threats faced the public health and medical science communities. Although the advances in medical science are formidable the increasing population of urban areas and the massive human mobility among world regions are factors that accelerate the spread and diffusion of old and new diseases [8,9,30,34]. As a result, we the research community strive to achieve predictive power to anticipate the unfolding of future outbreaks and evaluate associate risks [3,4,14,35]. The mitigation and containment of the next pandemic is crucially dependent upon our capacity to predict its diffusive pattern and optimize medical and mitigation policies [3,4,12,14,21,34]. In such containment processes, the ability to forecast how a disease might spread on the local and global level (as much accurately as we can now do for the weather) is essential for the identification and development of appropriate control strategies [3,14,21,34,35]. Along with infectious diseases, global health issues relates to access to resources, health infrastructures and a number of other determinants that shall be studied in a systemic perspective. Furthermore, a wide range of social contagion processes where the acquisition or increase of a behavior or cultural habit is related to social contacts between the uninfected population and the infected population. The contagion metaphor has already proven useful in understanding social behaviors such as tobacco, alcohol, and heroin use. In addition, it is becoming increasingly evident that social interconnectedness and opinion spreading through modern rapid means of communications may have important effects on the decisions taken by individuals with respect to ongoing health crisis; for instance the population adherence to vaccination campaign. The health and contagion exploratory has a particularly challenging undertaking, because it requires a truly interdisciplinary effort that combines and integrates knowledge, datasets, and techniques developed in epidemiology, medicine, physics, biology, computer science, and social and behavioral sciences. To reach this goal, knowledge as well as resources need to be accessed, shared and integrated among researchers working in this area. Feedback loops connecting the different areas of research are fundamental in allowing further improvement in a multidisciplinary field of research that otherwise would be hampered by disciplinary boundaries. Data availability as well as uniform and standardized approaches is an essential requirement. Analogously, quantitative modeling approaches should be made available to a large audience of users, including researchers, non-experts, policy makers, and health institutions [21,34].

By building on previous efforts and successes of projects such Epiwork it is possible to produce non-incremental advances in the capability of forecasting the spreading of contagion phenomena, relying on modeling and computational tools able to provide scenario forecast, risk assessment and containment measure. [3,14,35]. In this area the Epiwork project has already achieved major results in building a novel participatory platform for disease monitoring running or being tested in the wild in a dozen major European countries monitoring a population of about 250 million European citizens. Along with this platform a data-sharing infrastructure for epidemic data has been built and interfaced with a modeling platform that offers state of the art modeling tools. The Epiwork project has been involved in the effort of the international scientific community to fight the 2009 H1N1 pandemic. The project achieved several major achievements related to the H1N1 pandemic using computational models to produce projections of the spread of the ongoing H1N1 Flu epidemics [3–5,35]. The work done for the realistic modeling of the H1N1 pandemic has to be considered as a major breakthrough that has shown for the first time, in a real world situation, the potential of computational methods in providing anticipations and forecasts that can be used in the support of the policy making and public health decision-making processes [3–5]. Although we still face enormous scientific challenges in the development of an epidemic forecast European-wide infrastructure, the past successes points out that developing a global health exploratory to tackle those challenges is within reach of the FuturICT initiative.

Crime and conflicts exploratory

One of the EUs main objectives is to preserve and develop justice, freedom and security whilst addressing the increasingly complex security challenges. Over the last decade EU has witnessed a shift in security threats. The fight against terrorism and organized crime, the protection of the external European borders, has gained importance. Addressing these issues requires new and more effective strategies prompting citizen concern. There is a growing need to endow EU institutions, national governments with common instruments to support policy making, investigation and law enforcement. Phenomena like terrorism, organized crime, corruption, cyber-crime, financial fraud, drug and arms trafficking arouse not only serious social alarm and a huge economic damage. For this reason, we aim to build a Crime and Conflict Exploratory able to run massive data mining and large-scale computer simulations of social failures related to criminal activities. This Exploratory aims at having a relevant impact on society, since the analyzed phenomena are fast growing and widely spread [29]. It is therefore necessary to understand their dynamics, in order to foresee and eventually fight them. Hence the Crime and Conflict Exploratory will start with researches about the dimensions of crime, in its various forms. Some examples:

- Costs of Corruption. Suffering from serious flaws in estimation techniques, the analysis available need to be taken only as approximations [33]. Several estimates provide the upper or lower bounds, ranging from 2 to 5 per cent of global GDP, which amount is from 800 billion to 2 trillion of current US dollars. The Corruption Perceptions Index (CPI) calculated by the ONG Transparency Inter- national ranks countries according to the perception of corruption in the public sector, giving an idea of the dimension of the phenomenon worldwide.
- Cost of Crime. According to the United Nations Office for Drugs and Crime (UNODC), the annual turnover of transnational organized criminal activities such as drug trafficking, counterfeiting, illegal arms trade and the smuggling of immigrants is estimated at around 870 billion.
- Costs of Terrorism. It has been estimated that the cost of war terrorism in US since 9/11 is over 1 trillion of current US dollars. In general, the economic and fiscal costs of terrorism in rich country are low, relative to the economic costs of combatting terrorism [22,28]. Temporal trends show that from 1970 to 2007 there was a significant increase (over 85,000 discrete events).

These preliminary data show the huge impact of crime on social, legal end economic systems. The Crime and Conflict Exploratory will monitor, study and predict the evolution of these systems within and between different societies. The goal is to provide useful technological tools and support the future decision making procedures. The Exploratory will be built upon an innovative approach view and technological instruments. This approach needs a general scientific-grounded strategy, since global threats, as crime and corruption require global answers, able to coordinate national different policies. Expected outcomes, beyond a high impact on the fight against phenomena causing social alarms, are the improvement of the quality of policy/decision making, the increase of tax revenues and the reduction in the costs of crime and crime prevention. The Exploratory will consolidate, connect and innovate knowledge in the field, with the aim of offering stakeholders (from citizens to political institutions) new instruments, data and technologies for coping with severe social challenges. It will be focussed upon producing concrete policy recommendations for specific case studies. More in detail, the Crime and Conflict Exploratory will:

- Develop a framework for designing a set of policy options supported by theory and simulation. For example, in the area of crime, balancing repression versus protection of collaborators with the justice system; in the area of conflict, suggesting conflict resolution policies such as power sharing and autonomy;
- Model the multilevel dynamics of complex systems and make conditioned anticipation of trajectories of crime patterns, criminality and conflict;
- Understand the evolution and dynamics of criminal organizations and their relation with legal systems;
- Develop structures for integrating data on crime and conflict, exploring hypotheses such as the dependency of crime/conflict expansion on weak (vs. strong) official institutions, or that crime and conflict can mutually reinforce each other, where criminal organizations play an active role in funding and participating in the activities of rebels and/or governmental militias;
- Explore the interface between crime and conflict. This will involve understanding how the two might impact each other (e.g. given limited state resources, crimes may be easier to commit, and harder to deter, in states in conflict) and building on research that explores patterns of insurgency from the perspective of environmental criminology, how approaches developed to understand one might inform the other.

Social resilience exploratory

The scientific goal of the resilience laboratory would be to increase the degree to which science can improve the resilience of societies by using new and emergent technologies in thoughtful, informed and intelligent ways. As well the exploratory aims at developing models, theories and general understanding of how resilience operates in ICT-enabled and ICT-transformed social systems. In principle, the basic research builds foundations for making the most of newest technologies in the service of creating emergent social processes of high quality that enhance the potential of sustainable growth, and resilience of societies to threats.

The diagnostic aspect involves understanding and monitoring the dynamics of resilience in real economic, social and psychological processes. The task of monitoring resilience requires combining various types of data relevant for diagnosing resilience. These data come from official statistics, social surveys and social data archives. This type of data can be used for analysis of long time trends and also for providing context for analyzing the fast time scale. The fast time data can provide a daily running diagnostic of societies. The input to them will come from monitoring and content analysis of press, Internet, social networks, blogs and forum. Such already existing tools as Google trends can also be useful in this resect. The data coming from mobile devices as well as sensors can also be used for daily monitoring [11,16,27]. The resilience exploratory can also analyze culture and its changes with respect to resilience. Analysis of art and social narrations can provide useful information concerning the state of a society with respect to the potential for self-organization.

The prescriptive role involves design of resilient systems: How to coordinate bottom-up processes in a self-organizing manner, so that they will produce resilient social and economic dynamics. This would involve developing Plan B: a set of principles, strategies and software tools for rapid recovery in cases of systemic failure. The example of a practical function that already has been implemented is provided by the organization ICT for Peace, which provides communication, and information services for areas in which emergencies have occurred.

Social well-being exploratory

One of the scientific challenges in creating indices for social well-being, will be to measure the relevant factors, globally and in real-time, with sufficient accuracy. Measuring social well-being is even more difficult than measuring GDP. Currently, reliable official numbers for GDP are published only with a delay of many months. However, new ways of measuring GDP have recently been suggested. For example, it seems feasible to estimate GDP based on satellite pictures of global light emissions [26]. Such estimates are possible almost in real-time. Similarly, it has been shown that health-related indicators (such as the number of patients during flu pandemics) can be well estimated based on Google trends data.

Therefore, the vision of FuturICT is to make the different dimensions of social well-being globally measurable in real-time [25]. This could be done by mining freely available data on the Internet, by sentiment analysis of twitter feeds and blogs, or by use of sensor data of various kinds. Recent attempts to measure happiness and its variation in space, time and across social communities point the way for this [13,23]. At present, the greatest difficulty in measuring human capital (such as education) or social capital (such as cooperativeness and solidarity) is the calibration and validation of corresponding indices. It is envisaged to do this by comparison with (online) surveys and (web) experiments.

Social well-being can be defined along different subjective and objective dimensions, both in relation to the individuum and the society. Objective dimensions that can be quantified using different measures, include personal wealth and its distribution across the society, environmental conditions including pollution and ecological factors, or education and political integration. Many of these dimensions are also covered by other exploratories, e.g. for economics, environment and are, despite their impact on social well-being, not further discussed at this point. Other factors, such as well-being depleted by systematic stress or mental diseases, physical health problems etc. are covered in the Global Health exploratory.

This leaves a focus of the social well-being exploratory on those issues that are primarily related to (i) subjective factors and (ii) social factors. How can we learn about these? Traditionally, there are surveys both on the regional and the state level, for example the Bhutan Gross National Happiness, which are limited because of their small sample size, low time resolution, and costs involved. Therefore, our goal is to combine the available data with data from online communities, which can be obtained instantaneously at low cost. Further augmenting such data with other digital traces, such as data from mobile phone usage or search queries allows to reconstruct the changing "mind" of the human society.

Successful examples include the emotional well-being monitored from twitter data, or the emotional spreading in online communities. Such analyses include the direct reconstruction of e.g. the mood from written expressions or the social network, but also indirect insights into activities, emerging trends in social life, or hot topics of vital interest for a large portion of the population. The challenges, on the other hand, result from the underrepresentation of the aged population that does not use these channels of communication in the same manner as the younger one. Also, it needs a coherent approach to aggregate data about the emotional well-being from various sources.

Different projects, such as the European project CYBEREMOTIONS, have already developed pieces of such a monitoring platform of emotional expressions. The focus there was on collective effects, such as the spread of positive or negative emotions in special social communities, e.g. in chat rooms and forums. Other projects such as LIWC allow more insights into an individual's personality. Both strands of research need to be integrated in an exploratory of emotional well-being.

In order to study the impact of recent events, e.g. during an economic crisis, on social well-being, it is useful to combine such insights, which could reveal correlations of different activities or interests. While search queries, e.g. about health symptoms, have been already mentioned, search trends allow to find out about the relations between individual well-being and other activities. We note that, each of these tools existing so far, only highlights specific patterns of a subset of the whole population. Therefore, the effort is in linking such insights together and extending them into valid, quantifiable measures of well being.

Instead of reading user data in different Internet portals, another strand of research activities should focus on specific online tools to generate specific data by means of targeted measures. Known examples are Facebook applications to measure the IQ of users, their literacy, etc. Such approaches could be combined with surveys or online tests about other topics related to well-being, starting from diets to reading habits, media consumption, etc. Yet another level are "game with a purpose", which use individual affinities to gaming to find out about certain behavior, mental capabilities, or social skills which are related to social well-being.

In conclusion, an exploratory of social well-being has to comprise information from different sources. In addition to classical survey data, it should involve monitoring of online data from social networking sites or search engines and the development of targeted applications to find out about subjective factors and social integration. Pieces of such analyzing tools already exist. But they need to be further developed, validated and integrated into a coherent framework of monitoring. In addition to the technical and computational challenges involved, the scientific challenge is in defining quantitative measures related to social well-being at large.

3 Integration, validation and ethical issues

In the implementation and integration of exploratories in the development of the exploratory of society, there are some issues that may deserve attention beyond the strict technical and scientific roadmap, namely the quality control of the data generated, intellectual property rights and governance.

Quality control

There are several ways to make sure that the quality can be explicitly assessed:

– In keeping with the classical motto of garbage in, garbage out, perhaps the most important method would be to make sure that the exploratory proposals be carefully selected based on scientific principles. Instead, it would make sense to introduce at least two phases of exploratory construction: a first phase allowing already functioning data systems that meet specific quality criteria to receive funding, and a second phase expanding the list to those exploratories that need more time to live up to these goals. Possible criteria would for example be (i) explicit evidence proving the datas validity compared to existing datasets in the area in question, (ii) peer-reviewed publications based on the data. Whoever proposes an exploratory must be able to show that it relies on state-of-the-art knowledge that corresponds to the highest standards in the discipline in question. If peer-reviewed publications are still not available, it would be prudent to expose the proposals to external review conducted by substantive experts.

- Once the exploratory in question is up and running, ongoing quality checking should continue without interruption. In particular, it would be helpful if the exploratory includes a feedback mechanism of a wiki type that allows users to report mistakes.
- The first priority should be to make sure that the existing core of pilot projects is sound before major integration efforts are undertaken in order to link together the exploratories. Too fast expansion and integration would threaten to pollute the data quality within the system, especially if the exploratories are tightly coupled.

Intellectual property rights

Clearly, the data held by the exploratories cannot be the intellectual property of the FuturICT itself, but has to rely on permission to integrate the data in a larger system. In this context, it is essential that all data be open source. If this is not the case, the replicability could not be assured. Since any higher level of coordinated exploratories, and for that matter, the exploratories themselves, would constitute federated structures of data, ownership can be quite tricky to track down. Therefore, extensive safeguards to render the origin of the data explicit are needed. This could be introduced through logon sequences that require the user to cite the source in question, and a system that tracks the original publications as complex ingredients in the overall system.

Integration

Exploratories will feature a networked structure since they would make the best use of existing data and also allow those running them to focus on data integration rather than data documentation and editing. Integration both within and above the exploratories requires constant coordination involving the data-producing institutions. In particular, the more policy relevant the goals of the exploratories become, the more important it becomes to make sure that data are updated on a continuous basis. Failures to provide updates by single institutions may render entire layers of derived information useless.

4 Expected impact

The Exploratory of society aims at bringing a paradigm shift in the study, anticipation and control of complex societies. The goal is not simply to reproduce known features of societies and socio-technical systems, but to forecast trends in their future evolution in order to help preventing or mitigating the onset of crisis over the long term. The successful realization of this ambitious goal requires a pan-European approach, bringing together partners with different fields of expertise (and combining the resources of existing measurement infrastructures). The proposed outcome, beyond being merely practical in nature, can have a clear impact in the theoretical understanding of complex systems theory and the study of large techno-social systems. The change of focus from reproducing data to forecasting trends will contribute to the development of new general classes of models aimed at quantitative prediction, opening thus the possibility to new lines of research and placing those in the proper framework to be adopted as a mainstream objective of ICT. The development of society exploratories implies major advances in the characterization, knowledge and understanding of large scale techno-social systems extending from the scale of the single individual to entire populations, providing the design of a new generation of tools and computational models. Society exploratories are a truly interdisciplinary effort that is going to bridge different scientific communities and foster the interaction between complex systems science, social, behavioral and cognitive sciences, ICT and data gathering. The potential contributions to the attractiveness of the European Research Area apply more widely to scientific performance, technological development capacity, policy-making and interaction with the society. Furthermore, society exploratories will form the cornerstone of a consistent policymaking process informed by the integration of complex systems science in a socially intelligent approach to planning and crisis management.

The publication of this work was partially supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 284709, a Coordination and Support Action in the Information and Communication Technologies activity area ('FuturICT' FET Flagship Pilot Project).

References

1. D. Helbing, et al., Eur. Phys. J. Special Topics **214**, 11 (2012)
2. D. Kossman, et al., Eur. Phys. J. Special Topics **214**, 77 (2012)
3. D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, et al., BMC Medicine **7**, 45 (2009)
4. D. Balcan, V. Colizza, B. Goncalves, H. Hu, J.J. Ramasco, Vespignani A., Proc. Natl Acad. Sci. **106**, 21484 (2009)
5. P. Bajardi, C. Poletto, D. Balcan, H. Hu, B. Goncalves, et al., Emerging Health Threats **2**, e11 (2009)
6. D. Brockmann, L. Hufnagel, T. Geisel, Nature **439**, 462 (2006)
7. J.C. Bohorquez, S. Gourley, A.R. Dixon, M. Spagat, N.F. Johnson, Nature **462**, 911 (2009)
8. V. Colizza, A. Barrat, Barthelemy, A. Vespignani, Proc. Natl. Acad. Sci. USA **103**, 2015 (2006)
9. V. Colizza, A. Barrat, M. Barthelemy, A.J. Valleron, A. Vespignani, PLoS Med **4**, e13 (2007)
10. M. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, F. Menczer, Third IEEE International Conference on Social Computing, 192 (2011)
11. C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.F. Pinton, A. Vespignani, PLoS ONE **5**, e11596 (2010)
12. M.L. Ciofi degli Atti, S. Merler, C. Rizzo, M. Ajelli, M. Massari, et al., PLoS One **3**, e1790 (2008)
13. P.S. Dodds, C.M. Danforth, J. Happiness Studies **11**, 444 (2010)
14. D.L. Chao, M.E. Halloran, I.M. Longini, PLoS Comput Biol **6**, e1000656 (2010)
15. D.L. Chao, et al., Amer. J. Epidemiol. **173** 1121 (2011)
16. N.A. Christakis, J.H. Fowler, PLoS ONE **5**, e12948 (2010)
17. The science of civil war, The Economist printed edition, April 21st, 2012
18. D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, et al., Science **323**, 712 (2009)
19. M. Eichner, et al., BMC Infectious Diseases **7**, 17 (2007)
20. S. Eubank, H. Guclu, V.S.A. Kumar, M.V. Marathe, A. Srinivasan, et al., Nature **429**, 180 (2004)
21. N.M. Ferguson, D.A.T. Cummings, S. Cauchemez, S. Fraser, S. Riley, et al., Nature **437**, 209 (2005)
22. P. Fischer, T. Greitemeyer, A. Kastenmuller, D. Frey, S. Osswald, J. Exper. Social Psychol. **43**, 964 (2007)
23. S. Golder, M.W. Macy, Science **333**, 1878 (2011)

24. M.C. Gonzalez, C.A. Hidalgo, A.-L. Barabasi, Nature **453**, 779 (2008)
25. D. Helbing, S. Balietti, Eur. Phys. J. Special Topics **195**, 101 (2011)
26. J.V. Henderson, A. Storeygard, D.N. Weil, NBER Working Paper No. w15199 (2009)
27. L. Isella, et al., J. Theor. Biol. **271**, 166 (2011)
28. P. Keefer, N. Loayza, Terrorism Economic Development, and Political Openness (Cambridge University Press, Cambridge, 2008)
29. Y. Kim, Space-Time Measures of Crime Diffusion. In Liu, L., Eck, J. (eds.), Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems. Information Science Reference (2008), p. 140
30. S. Merler, M. Ajelli, Proc. Royal Soc. B: Biological Sciences **277**, 557 (2010)
31. J.-P. Onnela, et al., Proc. Natl. Acad. Sci. U.S.A. **104**, 7332 (2007)
32. J. Ratkiewicz, M. Conover, M. Meiss, B. Goncalves, S. Patil, A. Flammini, F. Menczer, WWW 2011, 249 (2011)
33. J. Svansson, J. Econom. Perspectives **19**, 19 (2005)
34. A. Vespignani, Science **325**, 425 (2009)
35. W. Van den Broeck, C. Gioannini, B. Goncalves, M. Quaggiotto, V. Colizza, A. Vespignani, BMC Infectious Diseases **11**, 37 (2011)