1. Objectives

ERACEP refers to the research theme "Emerging research areas" within the 'Coordination and Support Actions' (CSAs) of the ERC work programme for 2009. Its **main objective** is the identification of topically emerging research areas and to analyse to what extent the activities supported by the ERC cover and contribute to these research areas. Thus the project intends to investigate whether the basic mission of the ERC as formulated below

"Its main aim is to stimulate scientific excellence by supporting and encouraging the very best, truly creative scientists, scholars and engineers to be adventurous and take risks in their research." (ERC 2008)[1]

is reached in a satisfying way. Thereby the project also contributes to the question, whether the bottom-up approach of ERC funding schemes meets the expectation being able to support new lines of research with impact on the development of dynamic creative and innovative areas of research in Europe.

The **concept of ERACEP** basically combines two perspectives (cf. figure 1). The first perspective concerns the identification of topically emerging research areas with the assumption that excellent, truly creative research can be found in particular in these emerging areas. The second perspective takes the view of ERC funding and explores how ERC-funded research themes map to the identified emerging areas and which impact can be expected.

Various methods have been suggested for detecting early research trends, in particular Delphi surveys,[2] expert-based brain storming,[3] or expert conferences. Although many such activities are undertaken in a lot of countries[4], there is an obvious scarcity of really new topics. Standard "future" themes suggested refer to very broad areas such as biotechnology, nanotechnology, superconductivity, information technology and similar fields which are already relevant in the present situation, based on the assumption that these fields have a further potential of relevant growth. Most measures of research policy action, e.g. in the context of regional development or government funding of research, refer to these fields. However, only few suggestions are forwarded as to really new promising fields at a lower level of aggregation. Therefore, it is useful to think about **new approaches to identify alternative new fields** with relevance for the future development of technology and innovation.

---

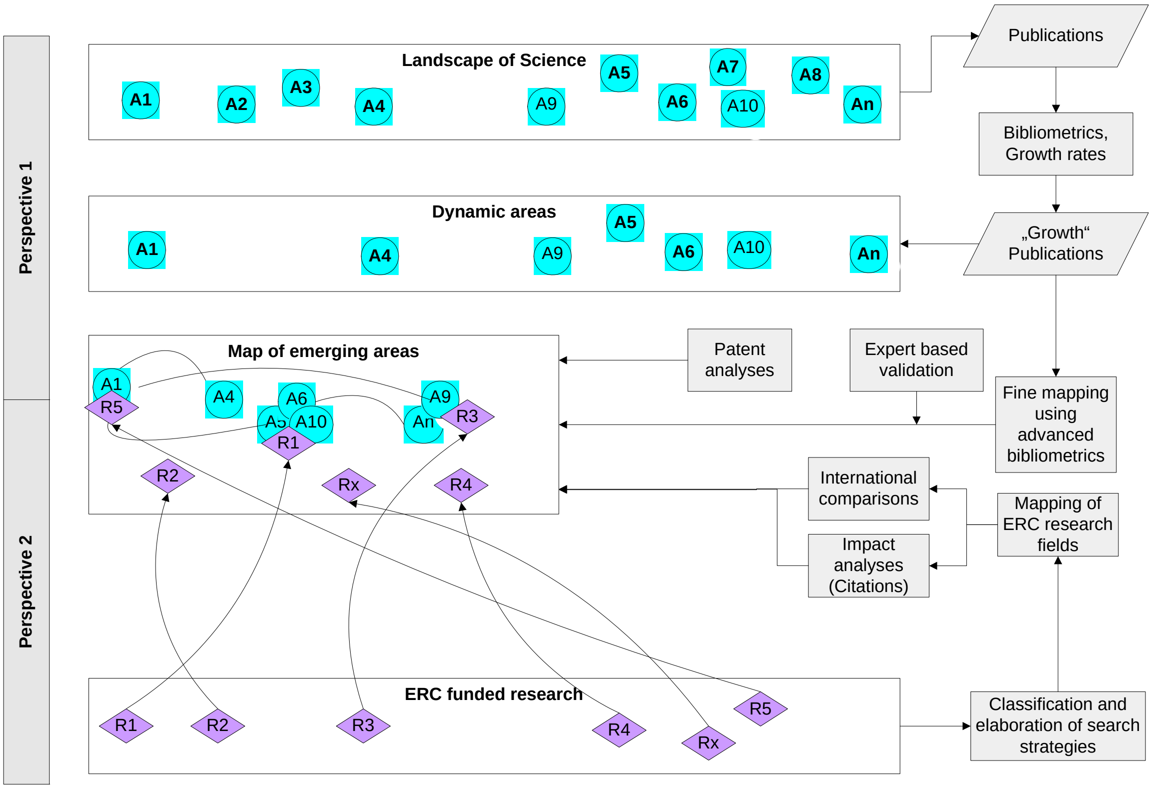
[1] European Research Council (2008): Mission, accessible through http://erc.europa.eu/index.cfm?fuseaction=page.display&topicID=12

[2] For instance, Cuhls, K./Grupp, H.(1998): Studie zur globalen Entwicklung von Wissenschaft und Technik,. Bonn : BMBF.

[3] For instance, Grupp, H. (1993): Technologie am Beginn des 21. Jahrhunderts, Heidelberg: Physica. In this study, nanotechnology was already discussed as relevant future field and the interfaces between traditional scientific disciplines were recognized as future drivers of research.

[4] The European Foresight Monitoring Network (EFMN) provides an overview of relevant activities: www.efmn.eu

Figure 1: Overview of the research approach (Ax: Research Area, Rx: ERC-funded research activity)



For the identification of emerging research fields, we suggest a bibliometric approach. We will use publications and not patents in order to observe new trends in an as early stage as possible. The specific advantage of a bibliometric search for emerging areas is its reference to objective data in contrast to the expert-based approaches cited above which inevitably imply a high level of subjectivity. Most bibliometric analyses with regard to emerging areas are based on different citation methods[5]. However, these approaches are linked to specific fields and need concrete entry points as prerequisite and are not appropriate for broad, open scans of all areas of science. Considering in particular the flexible bottom-up approach of ERC funding schemes, such open scans within the landscape of science are of particular importance for capturing and mapping ERC-funded research activities. Accordingly, in this project, we aim at an **open investigation in all fields of science** and we will use growth rates as selection criterion, as the distinct growth of a field indicates that a relevant number of authors consider it as challenging and promising; growth indicates a generally informal, invisible consensus within the scientific community.

[5] Takeda, Y., Kajikawa, Y. (2008): Optics: a bibliometric approach to detect emerging research domains and intellectual bases, in: Scientometrics, pp. 1-16. and Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K. (2008): Detecting emerging research fronts based on topological measures in citation networks of scientific publications, in: Technovation, 28 (11), pp. 758-775.

Several authors tried to scan publication and patent databases for detecting new areas through growths rates of documents referring to database codes (cf. Marmor et al. 1979).[6] However, it is difficult to assess growth rates of very small fields compared to large ones, and an enormous effort is required to eliminate artefacts. A further problem is the poor classification of most publication databases, so that the analyses refer to high levels of aggregation which do not allow for identifying the underlying developments in specific fields without additional steps of analysis. Furthermore, the analysis of large publication databases by exclusively quantitative methods implies the risk of the generation of artefacts, thus the detection of seemingly growing areas wherein the publications are accidentally linked by common keywords, but which have no common scientific basis.

We suggest an approach which takes the potential problems mentioned above into account. For this purpose, we will scan all areas of science in order to identify those with relevant growth in the last ten years. We will start with calculating growth rates of the standard areas of the Science Citation Index (SCI), the Social Science Citation Index (SSCI) and the Arts & Humanities Citation Index (A&HCI) and will transfer them into specific growth indices that take into account the different sizes of the areas. As the classifications of the SCI and SSCI are quite coarse, a further break-down will be made for areas with relevant growth by advanced bibliometric methods. The objective of this step is to identify the relevant sub-areas which sustain the growth of the area on the more aggregate level. The result of the identification of more specific sub-areas will be assessed by experts in order to check the coherence of the identified ones and thus to eliminate artefacts of the quantitative analysis. Special attention will be directed to the *social sciences* and *humanities*. Unlike in the sciences, emerging topics and fields in the social sciences and humanities tend to have three different types of genesis. Besides the rise of original research areas one can observe two other types here as well. One type corresponds to the specific interdisciplinarity of social sciences and humanities in the environment of which new areas can emerge. From the historical perspective, *urban studies* and the field of *bibliometrics* might serve as examples. The third type can be interpreted as repercussion to emerging topics in the sciences; these areas often evolve in parallel and in interaction with the corresponding science areas. Social aspects of HIV infections or of climate change are typical examples. Our approach will take these specific types into account by monitoring the evolution of research areas in the social sciences and humanities from the viewpoint of an inter- or crossdisciplinary environment and the environment of related emerging areas in the sciences.

The results of the bibliometric analysis will be a map of promising emerging areas distributed over all areas of science requiring excellent and creative research for their further development (cf. figure 1). By a supplementary patent analysis with regard to these areas in the natural, life and engineering sciences, the stage of their development will be characterised, in particular whether they are still in a very early stage rather or in a later stage where research primarily refers to specific high level applications.

These fields will be compared to the grants which were already supported by the ERC. By such a comparison it will be shown to what extent the ERC grants refer to these emerging fields or whether they refer to clearly distinct topics. In the latter case, their dynamics will be determined by additional bibliometric analyses as well. Furthermore, the activities of European scientists in comparison to American ones will be assessed by bibliometric means by publication shares and citation rates, i.e. in terms of quantity and impact.

In summary the project will have the following **impact on ERC's operation**:

[6] Marmor A C; Lawson W S; Terapane J F (1979): The Technology Assessment and Forecast Programme of the United States Patent and Trademark Office, in: World Patent Information, Vol. 1, No. 1, pp. 15-23.

- The ERC-funded research will be put into a systematic context with the worldwide research activities in emerging areas.
- It will be mapped within the landscape of science with regard to different dimensions of performance (scope, dynamics, impact, international uniqueness).
- In particular, the analysis will show whether the open calls of the ERC encourage the submission of proposals which really refer to dynamic areas of research and whether the finally selected ones fulfil this criterion. This information will allow to assess to what extent new scientific knowledge has been created from ERC-funded activities.
- Based on these findings, the appropriateness of the ERC procedures can be reflected in a systematic way so that the strategic objectives of the ERC support can be elaborated further on a rational basis.

2. Workplan

The workplan is structured into 4 parts (cf. figure 3). Each part consists of a number of workpackages with specific objectives.

Part 1 concerns the identification of emerging research fields. Within the landscape of science. Part 1 is broken down into the following 4 workpackages:

Workpackage 1 aims at the identification of dynamic areas within the landscape of science. For that purpose all fields of the databases SCI, SSCI, and A&HCI will be analysed with respect to their dynamics. For the determination of the growth rates, the number of publications in the last ten years in all fields of the selected databases will be determined. On this basis, the growth rates for the last five and ten years will be computed. In addition to the relative growth rates, the so called Sharp Ratio will be calculated (Fischer 2001: 271).[7] This is an index developed for the analysis of financial portfolios taking the size of the field into account. The use of this index proved to be useful in the context of bibliometrics, as the growth of small fields is generally higher than that of large fields due to the mathematical properties of small quantities.[8]

Workpackage 2 aims at a fine mapping of emerging fields within the identified growth areas of workpackage 1. For that purpose a further breakdown of areas with relevant growth into subareas will be made by advanced bibliometric analyses. The analysis of Emerging Research Areas will consist of two different conceptual approaches: (1) the structural topic analysis and (2) the dynamic analysis. For the first part, the full publication set of the Web of Science including the SCIE, SSCI and AHCI in the period 1998-2007 will be used in order to identify clusters and to determine their topics, for the second part, changes in time such as growth dynamics and convergence will be monitored on basis of separate annual volumes in the period 1998-2007.

The methods applied to the study is a combination of different bibliometrics techniques forming three basic methodological pillars, namely, bibliometric coupling among individual papers (Kessler 1963[9], Glänzel and Czerwon 1996[10]), cross-citation analysis among science fields (Janssens et al. 2009[11]) and text mining (Kostoff et al. 2001[12], Glenisson et al. 2005a[13],b[14], Janssens et al. 2006[15]) .

---

[7] Fischer, B.R. (2001): Performance Analyse in der Praxis. Performanzmaße, Attributionsanalyse, DVFA-Performance Presentation Standards, München/ Wien: R.Oldenbourg Verlag.

[8] Ulrich Schmoch; U; Wang, J.; Stoica, R. (2005): Research and Development in the Turkish and North African Landscape of Science. Report to the German Federal Ministry of Education and Research. Reiss, Thomas; Schmoch, Ulrich; Schubert, Torben; Rammer, Christian; Heneric, Oliver (2007): Aussichtsreiche Zukunftsfelder der Biotechnologie - Neue Ansätze der Technologievorausschau. Stuttgart: Fraunhofer IRB Verlag (ISI-Schriftenreihe Innovationspotenziale).

[9] M. M. Kessler, Bibliographic coupling between scientific papers, American Documentation, 14 (1963), 10-25

[10] W. Glänzel, H. J. Czerwon, A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level, Scientometrics, 37 (2), 1996, 195-221.

[11] F. Janssens, L. Zhang, W. Glänzel, Hybrid Clustering for Validation and Improvement of Subject-Classification Schemes. Information Processing and Management, 2009, to be published.

[12] Kostoff R.N., Toothman D.R., Eberhart H.J., Humenik J.A., Text mining using database tomography and bibliometrics: A review, Technological Forecasting and Social Change, 68 (3), 2001, 223-253.

[13] P. Glenisson, W. Glänzel, O. Persson, Combining full-text analysis and bibliometric indicators. A pilot study, Scientometrics, 63 (1), 2005, 163-180.

[14] P. Glenisson, W. Glänzel, F. Janssens, B. de Moor, Combining full-text and bibliometric information in mapping scientific disciplines, Information Processing and Management, 41, 2005, 1548-1572.

[15] F. Janssens, V. Tran Quoc, W. Glänzel, B. de Moor, Integration of textual content and link information for accurate clustering of science fields. In Proceedings of the I. International Conference on Multidisciplinary

Bibliometric coupling is based on the overlap of reference lists among different publications and gives a strong indication of thematic similarity of documents. The first and last methods will be used for the identification of clusters (cf. Janssens et al., 2009[11]), the other one serves as bibliometric tool of the evolutional analyses of clusters. The proposed “hybrid” approach allows to take the peculiarities of the *social sciences* and *humanities* into account. As known from experience and the relevant literature, citation analysis is a less efficient tool here. Hybrid cluster techniques allow to adjust the weight of their components so that the cluster analysis in the social sciences and humanities will focus on the textual aspect. This method will also make allowance for the links to related sciences fields and possible inter-or crossdisciplinary environment as described in Section 1.

Workpackage 2 will use the results of workpackage 1 as input. The structure of those subject categories, which have been found to grow at unexpectedly high rates or to potentially spawn new topics under their umbrella are further analysed. At the same time, the above-mentioned methods will be used to identify cross- and inter-disciplinary fields emerging from different established subject categories.

The best terms according to TF-IDF (Term frequency/Inverse document frequency) weighting scheme will serve to "label" the identified emerging clusters (see Janssens et al., 2009[11]). At the same time, core documents, that is, papers with strong links to numerous other papers (cf. Glänzel and Czerwon, 1996[10]) will be identified as potential representatives of clusters and thus to pinpoint clusters wherever possible and as possible link providers between different clusters.

The evolution of annual paper subsets within the clusters will be analysed to identify remarkable growth of literature. The evolution of distance among clusters within subject fields based on Euclidean similarity measures as well as of clusters among different fields based on cross-citation patterns will be analysed for possible divergence or convergence.

In workpackage 3 a validation of the emerging fields as determined in the previous workpackages will be made by expert assessment. As a result a subset of relevant fields will be extracted. For this validation step a selection of experts covering all fields of science is required.

In order to assess the stage of development of the identified emerging fields, additional patent analyses will be carried out for the areas of the natural, life and engineering sciences in workpackage 4. Thereby, in particular the application relevance of the fields can be determined.

Part 2 comprises a mapping of ERC-funded research to the emerging fields as identified in part 1. Part 2 is broken down into 2 workpackages. In workpackage 5 all ERC-funded grants will be classified to the system of emerging fields as identified in part 1. Within this workpackages also those ERC grants will be identified that refer to distinct topics which are not covered by the emerging fields.

In workpackage 6 those ERC-funded research activities which are outside the emerging fields will be characterised in more detail by bibliometric and patent analyses. In particular dynamics, impact and stage of development will be examined. In addition, international comparisons will be made allowing to assess how European research compares to research performed by American scientists.

Information Sciences & Technologies (InSciT2006). Current Research in Information Sciences and Technologies, Vol. I, Mérida, Spain, 2006, 615-619.

In **part 3** the methodological approaches for analysing the significance of ERC-funded research in relation to the landscape of emerging research areas will be compiled to a reusable toolbox (workpackage 7). Further, the empirical results of the first two parts will be analysed and conclusions on the impact of ERC funding on stimulating research in emerging fields will be drawn within workpackage 8. In a first step the project team will elaborate draft conclusions and recommendations on the role and impact of ERC-funded research. In a second step these conclusions and recommendations will be discussed and validated in an expert workshop.

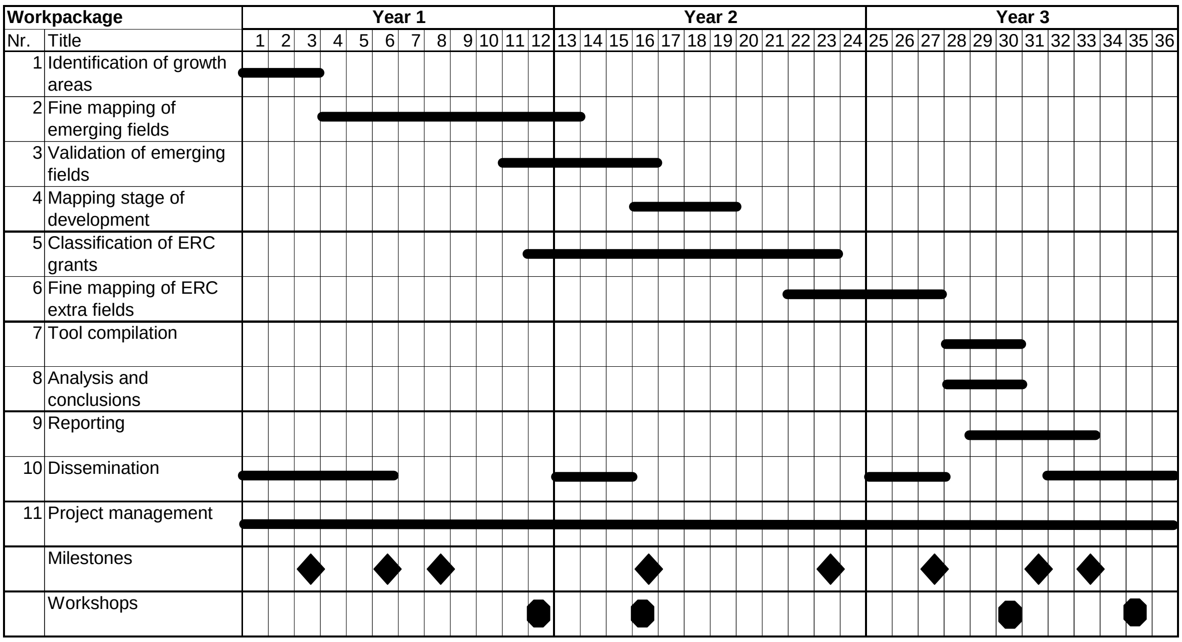
Part 4 of the project comprises reporting and dissemination of the results and is made up of the following workpackages.

In workpackage 9 the final report including an executive summary will be elaborated. In workpackage 10 the results will be discussed on a stakeholder workshop with representatives of the scientific community and European policy-makers. Finally, the project results will be disseminated and made publically available through a project website.

Project management tasks will be carried out within workpackage 11.

The timing of the different workpackages and their components is summarised in the following Gantt chart.

Figure 2: Timing of the different workpackages



The interdependencies between the individual components of the workplan are shown in the following Pert diagramme.

Figure 3: Graphical presentation of the ERACEP project