Cluster Life Cycles

Max-Peter Menzel* and Dirk Fornahl**

November 1, 2006

* University of Berne, Institute of Geography, Economic Geography and Regional Studies, Hallerstr. 12, CH-3012 Berne, Switzerland, Phone: +41-31-6318877, Fax: +41-31-6318511, email: menzel@giub.unibe.ch
** University of Karlsruhe, Institute for Economic Policy and Economic Research, Waldhornstr. 27, D-76128 Karlsruhe, Germany, Phone: +49-721-6086042, Fax +49-721-6088429, email: fornahl@iww.uni-karlsruhe.de
1. Introduction

In recent years, regional clusters have increasingly attracted the attention of politicians and researchers. Increasingly, the existence of one or several regional clusters is regarded as a prerequisite of regional prosperity. Different case studies describe clusters during different phases of their development (for example in Swann et al. 1998; Fornahl and Brenner 2003; Bresnahan et al. 2001; Menzel 2005), but many peculiarly focus on growing or functioning cluster. Grabher's (1993) story on the coal and steel district in the Ruhr in contrast gives a specific example of a declining cluster. Additionally, he shows how a cluster can hinder regional development. But also the now declining Ruhr area was at its time a dynamic, innovative and growing cluster that fostered firm formations and entrepreneurship. Another type of cluster is the emerging one. Only few studies exist on the emergence of a cluster, mainly because an emerging cluster is hard to detect and can only be described ex-post, like in Bresnahan et al. (2001). The few insights on cluster emergence led to the assumption that the processes responsible for the functioning of a cluster do not explain its emergence. The many failed attempts to create a cluster are evidence for this (Orsenigo 2001). For a more comprehensive understanding of regional clusters, it is necessary to explain their emergence, decline and their shift into new fields (Lorenzen 2005). The consideration of the different stages of a cluster helps to disclose the economic influence of clusters (Rocha 2004). This requires a dynamic or evolutionary approach.

There already exist different approaches which elaborate on the evolution of clusters (Pouder and St. John 1996; Tichy 2001; Audretsch and Feldman 1996b). But the existing approaches focus only on single facets of the different cluster stages. An approach that explains the growth of the cluster, change between the different stages and a shift of cluster to different technological fields is missing (Lorenzen 2005). The article at hand intends to fill this gap and presents a cycle model for regional clusters. The model distinguishes between different stages of economic development. A cluster can be in the phase of its emergence or can grow, decline as well as sustain its current state. Besides this distinction according to the quantitative development, the cluster stages are moreover characterized by a qualitative element, namely by the diversity of the available competencies in the cluster. Thus, the article follows the “resource-based view of the firm” (Winter 2003) which emphasizes the difference of firms regarding their competencies and networks. The firms (and institutions) of the cluster are the bearer of its technological diversity. The cluster is not seen as a homogeneous unity, but rather is described by the technological diversity of its protagonists and by the economic
utilisation of this diversity by the interaction of the different protagonists. Thus, the cluster constitutes itself from the interplay of its different elements.

The article proceeds as follows. In Section 2 we start with an examination of the different elements that comprise a local industrial cluster. Section 3 analyses the insights of research on industry life cycles and applies the identified characteristics to the cluster life cycle. Section 4 adds processes that are peculiar for the development of a local cluster and which result in a deviation of the cluster development from the industry development in general. After that the different stages of the cluster are examined and the peculiar characteristics of the stages are discussed (Section 5). Section 6 concludes and gives an outlook for future research.

2. Elements of Regional Clusters

We start with the question what regional clusters actually are. The discussion in Martin and Sunley (2003) points to the difficulty to answer this question. Porter (1998, p. 78) defines a cluster as follows: “Clusters are geographic concentrations of interconnected companies and institutions in a particular field”. Although this definition is very general, it contains the essential elements of a regional cluster. First, the cluster consists not only of firms, but also of a specific institutional environment. This institutional environment comprises beneficial institutions like cluster organizations, but also of research and educational institutions that are the basis for innovation networks and the formation of human resources.\(^1\) The firms of a cluster and their institutional environment cannot be seen separated from each other, since their respective development is connected closely to each other (Kenney and von Burg 1999; Maskell 2001). Institutions and firms are therefore the basic units of the cluster and the carriers of its competencies.

Second, the cluster concept implies that only certain firms and institutions belong to the cluster. Therefore, there exists an outer boundary of the cluster. The first delimitation refers to the fact that only firms and institutions “in a particular field” are relevant. A certain technological proximity, which represents the basis for various exchange processes and synergies, exists within a cluster. Firms and institutions, which make use of other technologies, are outside this “particular field”. Moreover, the firms and their institutional

\(^1\) Since the article builds on the cluster definition of Porter (1998), we apply the notion of the institution in Porter's sense. Social institutions like rules and conventions therefore are not part of it in this article.
environment are geographically concentrated and there is also a spatial boundary of the cluster besides the technological one. The spatial boundary delimits the firms of the cluster from firms in other clusters. Maskell (2001) points to the different cognition of knowledge asymmetries between firms within and between clusters. He argues that the low cognitive distance inside regional clusters enables the firms to cope with large knowledge asymmetries. At the same time, this low cognitive distance within the cluster is accompanied by a larger cognitive distance between clusters, even between different clusters of the same industry or technological field. Besides, the cognitive distance between firms of the same technological field, yet located in different regional clusters points to the fact that there exist regional specificities of the economic sectors. Therefore, the developments of the relevant industries have an influence on the development of the cluster without determining it because of the regional specificity of the cluster. But there is also the possibility that several clusters exists in one region that differ in their developmental logic. While the spatial boundary delimits the cluster from its industrial environment, the cluster is also detached from other parts of the production and innovation system within its spatial range by its technological focus. It only forms a part of the regional production system. Bathelt (2001) for example described the milieu within the emerging biotechnological industry in Boston that strongly differs from Saxenian's (1994) description of the milieu in the minicomputer industry in the same region.

The third and last aspect of the definition of Porter (1998) refers to the fact that the firms and institutions are “interconnected”. These connections refer to market exchange processes of goods and services, to imitation of behaviour as well as to cooperation, which require a high level of mutual trust, technological proximity as well as horizontal and vertical complementarities of technologically activities. These connections define of which economic sectors the cluster consists and which geographic extension it has (Porter 2003). These relations and exchange processes are not distributed evenly inside a cluster, but again form specific focal points (Menzel 2005). Depending on the level of abstraction a cluster can have several of these focal points, which in turn may form single clusters in themselves. For example, the storage media, the software and the semiconductor cluster in Silicon Valley are focal points (or sub-clusters) of the greater computer cluster. While the boundaries of a cluster are fluent, its centre is definable. Therefore, the cluster constitutes itself from a critical mass of thematically focused formal and informal exchange processes between firms and institutions, which take place in a spatially restricted area. We confine ourselves to those processes which take place within a cluster and which are the extraordinary features of
clustered firms in contrast to firms outside of regional clusters. Those factors which apply both to firms inside and outside regional clusters are left out of the analysis.

Concluding, the elements of a cluster are firms and institutions and their interconnections within a technological and spatial boundary. For a dynamic approach on clusters, the change of these elements has to be considered, too: how do the spatial and technological boundary of the cluster change, how does the interconnection change and at least, how do the firms and institutions change during the cluster evolution.

3. Industry Life Cycles and Regional Clusters

As discussed above, a cluster has two boundaries: a technological or industrial boundary that separates it from other firms and institutions within its spatial range, and a spatial boundary that delimits it from other firms and institutions in the same industry, yet in other places. The cluster thus is at the interface between industrial and local dynamics. The elaboration of this interface requires the elaboration of its constituting parts, namely industrial and localised development.

3.1. The Industry Life Cycle

The industry life cycle approach explains industrial change in analogy to the product life cycle. Like a product, also an industry follows cyclical development patterns. Klepper (1997, p. 148) distinguishes three different stages of an industry life cycle; embryonic, growing and mature:

In the initial, exploratory or embryonic stage, market volume is low, uncertainty is high, the product design is primitive, and unspecialized machinery is used to manufacture the product. In the second, intermediate or growth stage, output growth is high, the design of the product begins to stabilize, product innovation declines, and the production process becomes more refined as specialized machinery is substituted for labour. Entry slows and a shakeout of producers occurs. Stage three, the mature stage, corresponds to a mature market. Output growth slowly, entry declines further, market shares stabilize, innovations are less significant, and management, marketing, and manufacturing techniques become more refined.

In this model, there are few firms and employees in the embryonic stage, an increasing number in the growth stage that again declines in the mature stage. But neither the age, nor the quantitative development of firms and employees describe sufficiently the development of an industry. The dominant design (Abernathy and Utterback 1978; Suarez 2004) and the technological trajectory (Dosi 1988) point out that the development of industries and firms is
characterized very strongly not only by quantitative, but also by qualitative factors. During the emergence of a new technological trajectory alternative concepts compete with each other. Therefore, the technological field is very heterogeneous and the uncertainty over the future direction of the technology path is very high. Over time certain alternatives or development directions crystallize as the most promising ones, while other possibilities are rejected. By the cessation of possibilities that are considered to be inferior and by the fact that incumbent and newly formed firms orient themselves towards the most promising development approaches, the technology path increasingly focuses and the heterogeneity declines. The uncertainty about the further development of the industry is reduced due to the increasing stabilization and focusing of the technology path. This stage is accompanied by a growth of the appropriate industry (Suarez 2004). However, if the technological diversity cannot be maintained within the technological trajectory, the trajectory narrows increasingly. By such processes the firms, or the firms linked to this trajectory, can loose their ability for renewal and adaptation. The result is a lock-in that leads to a declining process in the corresponding industry. Therefore, there is a quantitative development of the industry described by number of firms, employees or turnover and a qualitative development that comprised the diversity of knowledge and competencies in the industry.

3.2. Quantitative and Qualitative Development of Clusters

Cluster development resembled the development of the industry life cycle in several aspects. As a rule, the stage is described by the age and the growth of the cluster in analogy to the product or industry life cycle (Enright 2003; Dalum et al. 2005). In these models the clusters develop like the industries with the consequence that age and growth stage are equated: Young clusters grow while old clusters are sustaining at best. This holds also for the qualitative development. Even if concepts like technological trajectory and dominant design cannot be assigned one to one to regional cluster, they nevertheless point to a connection between technological diversity and growth that can also be found in regional clusters. Grabher and Stark (1997) explain the sustainable growth of industrial districts in the so-called Third Italy with the variety of firms’ organisational forms that enables continual adaptation. Audretsch and Feldman (1996b) emphasize the significance of technological diversity for the development of regional clusters. The retention of diversity seems to be difficult especially in older clusters. Grabher (1993) puts forward a too low variety of organisational forms and technologies as one of the reasons for the negative development of the industry in the Ruhr district. In connection with this, Tichy (2001) points to the “cluster paradox”. On the one
hand, a narrow specialization of the cluster increases the possibility to utilize technological synergies between the firms. On the other hand, this strong similarity of the firms decreases the probability for more radical innovations, which would lead to a widening of a development path as well as to an increased ability of the cluster to adapt to changing external conditions. Therefore, a large diversity enables the cluster to adapt to changing environmental conditions or even to leave the previous development path. But a too large heterogeneity of the firms can also prevent an exploitation of synergies between the firms. Orsenigo (2001) describes this process with the failed biotechnology cluster in Lombardy. Their heterogeneity prevents firms to exploit technological synergies.

The diversity and variety of the existing and applied knowledge in the cluster is therefore a factor that has an effect on cluster development. However, the diversity of firms has to be examined in relation to the number of firms located in the cluster. Large clusters like the computer cluster in Silicon Valley consist of thousands of firms. Such clusters can dominate whole industries. But also the industrial districts of the Third Italy represent regional clusters albeit smaller and in a specific form. For example, the firms of the industrial district in Carpi are strongly specialised by their focus on woollens. In contrast to this, the Silicon Valley example shows a very large diversity of technologies and firms. Nevertheless, both cases are regarded as positive examples of regional clusters. The size and the technological diversity of clusters must therefore be in certain corresponding relations to each other. Large clusters can contain more technologies than small clusters and nevertheless generate sufficient synergies between firms because of their size. Smaller clusters must be focussed to be able to make use of the locational advantages of the existence of firms of the same industry. A small cluster that covers the same technological areas as a large cluster is therefore technologically more heterogeneous than the large cluster. This ratio of diversity to size is described with the terms focused and heterogeneous.

Concluding, regional clusters are distinguishable from one another by a quantitative and a qualitative dimension. The quantitative dimension describes the economic development of the cluster in terms of the number of active firms and employees. Because of the possible shift of the cluster into new industries, a description of the cluster by its development is more appropriate than a description by its age, as also old clusters can grow when they move into new fields. In analogy to the industry life cycle (Klepper 1997) we distinguish the cluster into the following stages of development: emergence with only few firms, growth with a growing
number of firms and employees, and declining with decreasing number of firms and employees. With the sustaining stage, we add a fourth stage to it that accounts for the fact that a cluster can be able to sustain itself on a high level of economic activity whose changes in number of firms and employees are of more cyclical than structural nature. In addition to this quantitative account, the qualitative dimension describes the cluster inherent heterogeneity of firms’ competencies available in the different stages. Figure 1 clarifies this connection between the two dimensions. During the emergence of the cluster, only few firms exist and the heterogeneity increases strongly because every new firm ventures in new technological areas of the cluster. In the growth phase the technology path focuses increasingly. The heterogeneity shrinks until the cluster has matured and a distinct development path has taken shape. However, if the cluster is focused too narrowly, it loses its ability of renewal and it declines.

![Cluster Life Cycle with respect to Developmental Dynamics and Heterogeneity](image)

This cluster life cycle is stylized and not deterministic. The heterogeneity of the competencies available in the cluster is, however, fundamental for its development. If its heterogeneity sinks, the cluster declines. If in turn heterogeneity rises, a new growth phase can result. Thus, the cluster can move back in the stage sequence, if its heterogeneity increases again.

4. The Systemic Dimension of Cluster Development

Beside the similarities and interconnections between cluster and industry development, there exist also some differences. The technological heterogeneity is one important aspect for the cluster life cycle. Bearers of the technological heterogeneity are firms and organisations. How
the firms utilise the diversity during the industry life cycle and generate new knowledge and
diversity has several implications for the cluster life cycle. Clusters represent the localisation
of a part of the production and innovation system of an industry. As such, they differ in
several aspects from diverse agglomerations or firms outside clusters. Baptista (2000) shows
that knowledge diffuses faster between clustered firms than between firms outside a cluster.
Rocha and Sternberg (2005) indicate that clusters positively affect the firm formation rate.
Especially Audretsch and Feldman (1996a) and Pouder and St. John (1996) point to the fact
that the development of clustered and non-clustered firms differ during different stages of
industry development. Additionally, under the same overarching market and technological
conditions, both clusters that grow and shrink can exist. Perhaps the best-known example of
this is the comparison of the booming computer industry in Silicon Valley with the shrinking
computer industry in the Boston region in the middle of the nineties of Saxenian (1994).
Additionally, the whole cluster can leave the old industry and move into new or related
industries. Klepper and Simmons (2000) for example show that the TV producer industry was
formed mainly from entrants from the radio industry. By their diversification into the TV
industry, the radio producers entered a new industry with a new life cycle. Thus, there is a
cluster inherent component that has an effect on growth and decline of a cluster independent
of the particular industry life cycle. This component is the utilisation of the diverse
competencies that differs between clustered and non-clustered firms.

The diversity of competencies in industries and clusters only takes effect if it is utilised by the
firms economically. Nothing else describes Granovetter (1973) with his winged notion of the
"strength of weak ties" in that networks work primarily by making the differences between
protagonists economically exploitable. The utilisation of the diversity of the cluster’s
competencies, compared to the development of the whole industry, depends on the
localisation of the innovation process, or as Malmberg and Maskell (2006) call it “localized
learning”. Localised learning, in their notion, depends on two factors. The first one is local
capabilities that are “some forms of knowledge creation and exchange that are still very much
rooted in the cultural, institutional, and social structures of particular places” (Malmberg and
Maskell 2006, p. 3). Local capabilities explain that economic activity is not evenly distributed
in space, because some kinds of knowledge are only available at particular places and
therefore are a kind of scarce resource. The other aspect of localised learning refers to the
knowledge exchange in spatial proximity than can take several ways. One is by direct
interaction. Firms collaborate and learn from each other. But there is no necessity of direct
interaction to learn. It is sufficient that firms monitor each other and see what they are doing. Firms assess the behaviour and results of their competitors’ activities and combine their ideas with their own knowledge. Additionally, firms can learn by social interaction of their employees.

The difference between localised learning and learning between actors at different locations lies not in the transaction cost or the efficiency of knowledge transfer. It is rather the kind of knowledge and information that is transferred and the different knowledge sources that actors have access to. Storper and Venables (2004) describe with the word “buzz” the informal, often diffuse but steady and pervading information stream within the region or a cluster. Grabher (2002) terms the same effect “noise”. Steady interaction, informal knowledge exchange, monitoring, and collaboration decreases the cognitive distance between actors. Malmberg and Maskell (2006, p. 5) describe this as follows:

All kinds of inter-firm learning are enhanced because spatial proximity provides them with an arsenal of instruments to obtain and understand subtle and complex information of possible relevance. The ability to benefit fully from knowledge heterogeneity among firms or individuals is, consequentially, closely related to spatial proximity through cognitive correspondence, but also by sharing a common history, identical jurisdictional order and equivalent factor costs.

Accordingly, firms in a cluster can learn with and from more heterogeneous actors in larger technological distance within a cluster than outside a cluster.² Saxenian (1994) shows in her comparative study of the computer industry in Boston and Silicon Valley that (at the time of the study) the advantage of the firms in Silicon Valley lays in the constitution of the local networks that enables the firms to make use of the diversity of the available competencies. Longhi (1999) describes, at the example of the technology park in Sophia Antipolis, the transformation of a “satellite platform district” (Markusen 1996) into an operating cluster. The original branches were not connected to each other. Only by spinoff processes they were embedded in regional network structures. A large diversity of competencies existed all the time, based on the satellite platform firms. But this diversity could only be exploited due to the different network connections of the spinoffs. The newly available diversity of competencies also has given new growth impulses to the existing firms. Therefore, not only the diversity of competencies is of importance for the cluster life cycle model, but also how these specific competencies can be utilised by other firms and, in turn, how single firms

² An important point to mention here, and to come back later to, is that for learning from heterogeneous actors these actors even must exist in the cluster.
influence the processes of innovation of other firms. These effects, which arise by the mutual influence of different elements, are called “systemic effects” in the following. The systemic effects accounts for the fact that firms are part of a complex production and innovation system. They are connected to other firms and institutions of the cluster by various exchange relations and mutual interdependences. Therefore, firms have also an effect on other firms and institutions of the cluster and therefore on the complete cluster system.

These systemic effects, however, are not only present for the qualitative dimensions. The quantitative dimension, measured as the number of firms and employee, has an effect which goes beyond the contribution to cluster size. The size of the cluster affects, for example, how the cluster is perceived. Larger clusters are more likely to be perceived by various economic agents while smaller clusters can be ignored more easily. A stronger perception can lead to better supports from the political side. Prevezer (1998) describes, for example, that an incubator was set up in a biotechnology cluster in North-Carolina after the cluster already contained over one hundred firms. Feldman et al. (2005) show for the capital region around Washington, D.C. that the venture capital has not initiated the development of the cluster but attained meaning only in its later development. In addition, the regional universities have offered cluster-related teaching curricula only after the emergence of the cluster in the region. Longhi (1999) gives a similar example of Sofia Antipolis. A university with a cluster-related education was set up there only after the first growth phase of the cluster. These cases show how important the perception of a cluster is and that the perception depends on the size of the cluster. Another quantitative systemic effect is the ability of the firms in the cluster for joint and collective actions to put through their concerns. The larger the number of firms and employees in a cluster, the larger is the collective influence these firms have and the more likely it is that they succeed in satisfying their specific needs. An example is given by Saxenian (1994): in the forties a large portion of the national defence spending of the USA went to companies at the East coast. This relationship deteriorated further when firms of the west coast lost some orders. In 1943 an association was formed by Californian companies of the electrical industry with the aim “to promote their industry, particularly by lobbying for a share of defence contracts that were going to eastern companies” (Saxenian 1994, p. 21).

The dimensions of a cluster are summarized in Figure 2. The quantitative dimension consists of the number of firms and employees while the qualitative dimension integrates their specific technological competencies in the model. Both dimensions do not only have a direct effect but
also influences other parts of the cluster. This aspect is taken into account by the systemic dimension. The systemic quantitative dimension describes the perception of the cluster by external agents as well as the ability of the firms and institutions for collective actions. In addition, the systemic qualitative dimension describes the effect of the competencies of the individual firms and institutions on the innovation system.

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct</td>
<td>number of firms and institutions</td>
</tr>
<tr>
<td></td>
<td>number of employees</td>
</tr>
<tr>
<td></td>
<td>variety and diversity of competencies and organizational forms</td>
</tr>
<tr>
<td></td>
<td>heterogeneity vs. focussing</td>
</tr>
<tr>
<td>systemic</td>
<td>perception of the cluster</td>
</tr>
<tr>
<td></td>
<td>capabilities for collective actions</td>
</tr>
<tr>
<td></td>
<td>exploitation of synergies</td>
</tr>
<tr>
<td></td>
<td>networks and value chains</td>
</tr>
</tbody>
</table>

Figure 2: Dimensions of the Cluster

5. Stages of the Cluster Life Cycle

The differences in learning between clustered and non-clustered firms result in their different qualitative development. Firms do not remain static in a particular field, but move through the technological space during their development (Stuart and Podolny 1996). The extent of this qualitative movement is limited by past experiences and path dependencies (Dosi 1988). The direction of the movement depends on the source, where knowledge for the learning process emanates. When firms learn, they combine their existing knowledge and their resources with the new knowledge. In doing so, they adjust their knowledge to the sender of the knowledge and move towards its direction (Denzau and North 1994). While firms adjust to all other of their collaboration partners independently of their location, the specific character of local adjustment is that it takes place over larger technological distances and a larger technological variety because of the small cognitive distance between firms in a cluster compared with non-clustered firms (Maskell 2001).

When firms move rather in the direction of other firms within the cluster, especially towards success stories and the cluster’s dominant design (Abernathy and Utterback 1978), everything else being equal, technological distances between firms and, thereby, the technological diversity and heterogeneity of the cluster decrease. Therefore, the evolution of clusters is marked by a continual bridging of technological distances between its firms and respective
adjustment of the firms towards each other. But for the firms to be innovative, an optimal technological distance is most appropriate (Wuyts et al. 2005). A too small distance hinders the actors to create novelty while a too large distance hinders the actors to communicate. Accordingly, the cluster growth when the firms have optimal technological distance to each other and declines when the distance is too small or too large.

These effects have the following impact on the cluster life cycle: During the emergence of the cluster, the smaller cognitive distance between the clustered firms enables them to generate earlier exploitable technological distances and utilise more synergies compared to the non-clustered firms. During the growth of the cluster, the firms further adjust to each other and generate optimal technological distances. But the implicit tendency to adjust to each other would lead to a suboptimal and to small technological distance that only could be balanced in the sustaining cluster by steady implementation of external knowledge. This balance effect by external knowledge is not necessary for non-clustered firms as they always rely on external knowledge sources. If the clustered firms cannot achieve this balance, they develop worse than the non-clustered firms and the cluster declines.

It can be difficult to assign a cluster to a concrete stage if it is in a transition between two stages. A cluster consists of many different protagonists who develop differently. Only the sum of all these independent developments represents the development of the cluster. However, this heterogeneity also leads to the result that the cluster does not develop evenly and as a whole. Parts of the cluster can stay in an earlier stage while others already are in a later one. We assume, however, that the cluster as a whole will enter in a new stage after this transitional phase.

Reasons for the transitions between the stages can be external factors like developments of markets, but in this paper we concentrate on the dynamics within the cluster to explain transition between the stages. In contrast to the external factors, diversity is something that can be influenced by regional protagonists. Thus, the crucial aspect for the transition between the stages in this article is the diversity of competencies, as direct qualitative dimension, and the exploitation of diversity, as systemic qualitative dimension. This diversity can be maintained or increased by firm formations, diversifying firms as well as access to knowledge sources outside the cluster. The aim should always be the retention of a certain level of heterogeneity. Focussing and specialization are only relevant for the emerging cluster in order
to generate first growth potentials. The different cluster stages are described based on their specific characteristics in these dimensions in the following. Figure 3 gives a summary of the characteristics of the stages.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Direct</th>
<th>Systemic</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emerging</td>
<td>few firms and employees</td>
<td>cluster hardly perceivable</td>
<td>quite heterogeneous</td>
</tr>
<tr>
<td></td>
<td>direct</td>
<td>systemic</td>
<td>scarce possibilities to form networks</td>
</tr>
<tr>
<td>Growing</td>
<td>growing number of firms and employment</td>
<td>growing perception</td>
<td>growth of absolute diversity, decrease of heterogeneity</td>
</tr>
<tr>
<td></td>
<td>direct</td>
<td>systemic</td>
<td>open and flexible networks contribute to exploit diversity of competencies</td>
</tr>
<tr>
<td>Sustaining</td>
<td>stagnating number of firms and employment</td>
<td>the cluster shapes the region</td>
<td>homogeneous or focussed competencies, strong bias of the regional economy towards the cluster</td>
</tr>
<tr>
<td></td>
<td>direct</td>
<td>systemic</td>
<td>open networks contribute to utilize existing synergies and external knowledge</td>
</tr>
<tr>
<td>Declining</td>
<td>decline of the number of firms and employment</td>
<td>negative sentiments with respect to the cluster</td>
<td>low diversity and strong focus on a narrow technological trajectory</td>
</tr>
<tr>
<td></td>
<td>systemic</td>
<td></td>
<td>closed networks result in an insufficient adaptability of the cluster.</td>
</tr>
</tbody>
</table>

Figure 3: Distinguishing Features of the Cluster Stages

**Emerging clusters.** It is difficult to exactly define the phase in which a cluster arises. The main reason for this is that the emerging cluster is not a cluster actually. A spatial concentration of firms of a specific economic sector hardly exists. A specific institutional environment of the cluster does not exist, too. The corresponding economic activity in regions with an emerging cluster can almost not be distinguished from regions without an emerging cluster. Therefore, it is very probable that an emerging cluster is not perceived at all. The stage of emergence can often only be described ex post (for example in Bresnahan et al. 2001).

The few firms of the cluster are scattered technologically over wide areas. Although these firms can already represent the future technological orientation of the cluster, the cluster is
very heterogeneous due to the low number of firms. This heterogeneity aggravates exchange processes between the firms. Therefore, networks and customer-supplier relations only exist in parts of the emerging cluster. Co-operations between firms within the emerging cluster take place only partially because of the high variability of the few firms. For that reason, mainly synergies between firms and the scientific infrastructure are relevant in this stage (Shohet 1998). There are, however, two aspects in which a region with an emerging cluster differs from regions without clusters. Firstly, one or several firms exist which offer a lasting vision for a regional technology path. Secondly, certain conditions are given in the region, for example a strong scientific base or political support, by which the emerging cluster has a potential to reach a critical mass.

There are two different possibilities for the end of this phase. The first possibility is the transition into a growing cluster when the firms manage to exploit synergies between them and create network externalities. This can occur due to shifts of existing firms towards the centre of the cluster or a closure of competence gaps by further firm formations. The growth of the first formations and subsequent spin-off processes, often from one single firm, are regarded as responsible for the first growth of a cluster (Klepper 2001). One explanation for this is that the superior routines of the successful firms are inherited to its spin-offs that thus also grow above average. But it is not only the inheritance of successful routines that makes spin-offs crucial for the emerging cluster to reach a critical mass. When the spin-offs stem from the same origin, they are not only technologically close, but also connected by various social networks that further decrease the cognitive distance. The same origin enables collaboration and synergies between the firms that would not count for firms from different origins. Because of their ability to generate synergies, “incubator networks” (Menzel 2005) that are formed by firms that base upon the same origin are crucial for the emergent cluster to reach a critical mass to reach the growing stage. Although an endogenous and continual transition from the emerging stage to the growing stage is possible, mainly during the growth of markets in which the incumbent firms are well positioned, often the crucial push stems from single events and a sudden change in some exogenous factors. In the Capitol region around Washington, D.C. (USA), for example, a bundle of events, e.g. new regulations and government intervention, led to the growth and take off of the cluster (Feldman 2001). Also in Sofia Antipolis a changing economic environment caused the restructuring of the incumbent firms, including dismissals of highly qualified employees, that in the end led to the transition
from a so-called “satellite platform” (Markusen 1996) to a growing RIC in software and telecommunications (Longhi 1999).

The second possibility to end the stage of emergence is when the emerging cluster loses its potential to become a functioning one. This could happen if the possibilities for exploiting synergies between the firms vanish. Decisive for this are two reasons. One is the loss of the common focal point of the emerging cluster. The firms develop in different technological directions and the technological distance between them extends. Orsenigo (2001), for example, describes an Italian case in which the biotechnology firms in the analysed region in fact tended to cluster, but this (emerging) cluster failed to reach a critical mass because, among other reasons, of its heterogeneity. The second reason is the disappearance of existing firms from the emerging cluster. These ‘lost’ firms leave competence gaps in the emerging cluster and this also diminishes possibilities for regional co-operation. In the end the firms of the formerly emerging cluster might completely disappear.

**Growing clusters.** A strong increase in employment resulting from the strong growth of incumbent firms and a high number of new firm formations indicates a growing cluster. Unlike in the case of the emerging cluster the boundaries are definable in the growing cluster. Both the incumbent firms and the firm formations orient themselves at the growth centres of the cluster and follow them. This narrows the boundaries of the cluster and the cluster becomes more focussed. The growing density of firms and institutions within the boundaries of the cluster increasingly creates possibilities for innovation networks and customer-supplier relations. The continuous arising of new potential network partners avoids an isolation of single networks. An economic utilization of these possibilities leads to further growth of the firms. The growth of the cluster generates a milieu, which has a positive effect on existing firms and firm formations. Examples are the early semiconductor industry in Silicon Valley (Saxenian 1994) or biotechnology in Boston (Bathelt 2001).

The growing stage ends when the quantitative development of cluster’s firms adjust to the development of the rest of the respective industry (Pouder and St. John 1996) and the cluster gets, at least for a short time, to the sustaining stage. The main reason for this adjustment is the exploitation of the diversity and a more focussed orientation of the cluster that led to a distinct manifestation of a kind of “dominant cluster design”. An example for this is Detroit.
After the growth stage with extensive firm formations, the finally construct with Chrysler, Ford and General Motors emerged in the cluster (Klepper 2001).

Sustaining clusters. The sustaining cluster is a deviation in the typology. This stage describes a kind of equilibrium state. A sustaining cluster shows neither large growth nor a remarkable decrease in the number of firms or employees. Fluctuations are rather of a more cyclical than structural nature. Nevertheless, the transition from a growing to a sustaining cluster can be accompanied by a decreasing number of firms (Klepper 2001). The various competencies of the firms are exploited by dense and established networks. The connections of the firms of the cluster to the outside firms and institutions bring new knowledge inside the cluster and keep the networks open. During its development, the cluster has shaped its regional environment. The development of the region is even equated with the development of the cluster if one cluster is very dominating. The cluster is capable of steering relevant developments independently (Bresnahan et al. 2001). Examples are car production in Baden-Württemberg or the industrial districts of the Third Italy.

There are two effects that end the sustaining stage. The first follows the cluster cycle. A decreasing diversity and a too narrowly focussed cluster in an exhausted trajectory leads to the decline of the cluster. The second possibility is a step back in the cluster life cycle and a new growth phase. This new growth can result from the generation of new diversity that is also accompanied with the entering of new markets, like for example in the Silicon Valley of the 80s. A threatening decline of the semiconductor industry due to increasing rigidity of the large chipmakers and fierce competition from Japanese manufacturers resulted in a new wave of firm formations in the form of spin-offs from engineers discontented with the prevailing conditions. These new firms extended the variety of the cluster and led to new growth (Saxenian 1990). Yet, the altering and renewal of the development path often takes place in the cause of a substantial crisis and therefore follows subsequently after a stage of decline (Martin and Sunley 2006; Meyer-Stamer 1998).

Declining clusters. A declining cluster is defined by a decrease in the number of firms and employees due to firm exits, mergers and rationalizations. Hardly any new firms are set up at the same time. A region containing a shrinking cluster is labelled by a strong cluster-oriented bias of economic activities. This bias works by a specific knowledge base, highly qualified and specialised employees and firms strongly focussed on specific markets and technologies.
The competencies of such a cluster are contained only in a few firms. Despite the decline, competitive pressure can lead to high innovation rates (e.g. in Grabher 1993, Graham 1956). These innovations, however, arise within the existing and exhausted development path. Due to the former success, the cluster is aligned with the further development of the formerly successful development path. A reason for this lies in the long existing, closed and homogeneous networks in which uniform knowledge flows (Granovetter 1973). A declining cluster has therefore lost the ability to sustain its diversity of competencies and to adjust to changing environmental conditions as well as the potential for an independent renewal.

There are three possibilities to end the declining stage of the cluster. The first, again, follows the cluster cycle and the cluster simply diminishes. The second possibility is the altering and renewal of the existent development path through the implementation of new and different technologies like in the accordion cluster in Marche that transformed from traditional towards electronic music instruments (Tappi 2005). The third possibility to end the declining stage is a shift towards completely different fields. This shift took place in the declining coal and steel industry of the Ruhr area towards environmental technologies. The firms of the steel and coal complex acquired the respective competencies during the correction of environmental damages that even originated from that industry (Grabher 1993).

6. Conclusion

The paper presented a cluster life cycle model to explain the evolution of regional clusters. We started with the findings that clusters show and move through different developmental stages and that these developmental stages can differ from the development of the respective industry. The cluster has a quantitative and a qualitative dimension and both of them have a systemic dimension. This systemic dimension accounts for the fact that firms and organisations are in a cluster are influenced are part of a complex production and innovation system and by various interconnections both influence, and are influenced by, other firms and organisations. Additionally, the systemic dimension describes the innovation and learning processes of firms that are crucial for the evolution of the cluster which makes the cluster distinct from the general industry development. These quantitative and qualitative dimensions show specific characteristics in the different stages of the cluster life cycle. Our hypotheses on these characteristics can be tested by empirical data. Furthermore, it can be analysed how peculiar processes are influenced by the stages of the cluster and how these processes influence the cluster development and the discussed dimensions. Such a process could be firm
formation for example. The influence of firm formations on the cluster in our model does not only confine itself to a contribution to the growth in the number of firms or employees. Since firm formations are embedded in different networks and exchange relations and therefore have an effect on other protagonists, they influence the complete production and innovation system. The characteristics of the production and innovation system are, however, dependent on the stage of cluster development. The influence of firm formations on the cluster is therefore different in the different stages of development. In turn, the characteristics of the cluster also have an effect on the kind of firm formations which it generates. This influence again depends on the stage of the cluster and has an effect both on the number of firm formations and on the technological areas in which they are founded.

But a lot of research still has to be done, including deeper theoretical analyses of the underlying processes and closer consideration of the external environment, as well as more empirical studies to test the model. The life cycle model consists of four dimensions that each requires distinct empirical methods and data. Köhler and Otto (2006) for example, tested the direct quantitative effect of firm formations and cluster development with quantitative data on firm survival rate, number of startups and number of closures with respect to employment growth on different clusters at different stages. Appropriate data to test for the development of the direct qualitative dimension would be patent data. The change of classification of patents could give information on the development of the diversity of the cluster over time. More difficult to measure are the systemic effects as their influence becomes most probably apparent only with in deep case studies. There still exists a lot of literature that independently describes the different dimensions of the cluster. The task at hand is to combine different kinds of studies within a single framework. Therefore, further empirical work on the cluster life cycle model would require both quantitative and qualitative research. The necessity to apply both types of research shows that neither case studies, nor econometric analyses can contribute alone to understand the evolution of clusters and indicates the complexity of the evolution of regional clusters. The cluster life cycle model could contribute to order this complexity in a meaningful way.
References


