**Pros and Cons of Tailor-made Data Solutions and Income Generating Activities to NSI activity**

– The case of Statistics Denmark

Mikael Skovbo, Statistics Denmark, mik@dst.dk

Dagmar Thomsen, Statistics Denmark, dht@dst.dk

**Abstract**

*Within the vision of being Denmark’s key provider of data that informs society, Statistics Denmark’s common strategy 2018-2022 defines a key mission to maximize its relevance and make sure that data ‘gets out there’ thus ensuring that publicly collected data is put to its maximum use. This paper makes the case that while official statistics is a key tool to this end, the scale up of tailor-made data solutions is an essential element to reaching the vision. Because tailor-made solutions ensure relevance and maximum use of collected data - but it in-fact also feeds back into official statistics strengthening it.*

*During the last year Statistics Denmark has had a turnover of approximately 100 million dkk on tailor-made data - translating into approximately 89 man years. We have serviced an average of 750 distinct customers, ranging from public to private companies - with data solutions that would not otherwise have been produced. To boost funded tailor-made data solutions Statistics Denmark is in the process of developing a new Strategy.*

*Adding to the increase of NSI relevance and use of data, the major gain entails feedback to the official statistical production strengthening it fundamentally. This feedback takes the form of new and/or improved variables, software developments, methodological improvements and closer as well as broader liaison with customers and stakeholders. All of which would also not otherwise have been the case. This article discusses these gains and the challenges posed by engaging in Income Generating Activities, and makes the case that tailor-made data solutions are a significant gain to, and could be seen as part of, the core business of NSIs.*

**Keywords:** Income Generating Activities, Tailor-made Statistics.

1. **Introduction**

By 1966 the Danish Minister of Economy at the time, Ivar Nørgaard, announced a new act introducing Statistics Denmark to the concept of Income Generating Activities in the statistical production. The purpose was mainly to ensure that the data and competencies developed in the official statistical production was put to a maximum use. That way society could be served with tailor-made data solution – while these activities would be financially self-reliant.

After the first introduction of the act in 1966, data has changed in character, registers have become more effective, and IT solutions allow for a large variety of statistical products – of which not all can become part of an official statistical production. Today microdata can be widely crossed for individual needs, using identification keys, providing substantial evidence for policy- and decision-making as well as research.

The law governing Statistics Denmark states that it is the Board of Governors that decides which statistics to produce as part of the public program – balancing amongst others the response burden, general usefulness of data and law requirements. In turn if the data allows it, what is not made by the public statistical program can be produced against private financing. To make use of this option the main requirement in the law is that the services to customers must be a *direct continuation of the core mandate of Statistics Denmark*.

On the mandate for Income Generating Activities provided by the act, four types of services are delivered: 1) Tailor-made aggregate/tabular data as requested by customers, 2) Microdata for research and analysis purposes and 3) user-defined surveys, sample development and optimization and finally 4) international capacity building and consulting.

Today IGA constitute 23 percent of Statistics Denmark’s Income, and it yearly employs a minimum of 89 man years directly. The magnitude of the activities in terms of tasks produced, customers served and direct turnover demand a strong reflection on the activities’ role, and their benefits and challenged for Statistics Denmark.

For this purpose in this paper we aim to answer the question *‘to what extent do the main benefits of privately financed Income Generating Activities outweigh the threats of the same and support these as a central element in the business of a National Statistical Organisation’.*

To answer this question we centre our analysis along the two main arguments that drive the argument for Income Generating Activities namely 1) putting data to its maximum use and 2) allowing for financial self-reliance and sustainability. Adding to these two key objectives we will briefly analyse a few subsidiary effects for - god or bad.

Finally we reflect on the challenges we face currently and how they are handled in our new strategy for Income Generating Activities.

1. **Putting data to its maximum use**

Income Generating Activities are a seemingly a key element for Statistics Denmark in ensuring that the data collected and produced is put to a maximum use. This argument has three central aspects.

1. The customized data produced serve a market that would not otherwise have been served, thus putting data to more use
2. The products produced as Income Generating Activities also provide methodological improvements and developments that feed back into the official statistical production fundamentally strengthening it and
3. The type of partnerships and stakeholders that Statistics Denmark’s works with under the parachute of Income Generating Activities ensure closer relations to users and create a better and broader understanding of user needs - ensuring a strong feedback link to the official statistical production

Below we analyse each of the arguments.

*2.1. Services and data that would never have been delivered*

The most important reason for engaging in privately financed customized data services, whether microdata or aggregated data solutions, is the fact that the statistical data that Statistics Denmark possesses is put to its maximum use.

Every year more than 750 individual customers are served products that have value to them. These customers receive more than 3000 individual services that would not have been produced otherwise.

The main reason for this is that Statistics Denmark would not have the financial capacity and subsequently manpower under the official statistical production to deliver these services. So in fact, more than 750 customers, and 3000 individual services are the added value of the Income Generating Activities.

Let’s examine the customers. The customers for tailor made solutions are primarily state institutions and universities. In 2018 Universities and Research Institutions accounted for approximately 43 pct. of the turnover, while ministries, municipalities and other public sector institutions accounted for 38 pct. While private companies only represent 14% of turnover – in numbers they represent 52% of all customers.

**Table 1. Customers in numbers**

|  |  |  |
| --- | --- | --- |
| Customer group | Number | Pct. |
| Research and Education | 92 | 12% |
| Civil Society Organisations | 64 | 8% |
| Private Companies  | 390 | 52% |
| State, Regions and Municipalities | 151 | 20% |
| Foreign customers | 57 | 8% |
| Total | 754 | 100% |

*Source: CRM, Statistics Denmark*

The turnover development is a result of state and research institutions’ increased need for tailor made data solutions for research, policy development and decision making.

These 52% and specifically 390 companies are dependent of the data provided by Statistics Denmark. Many of them contribute to the statistics that we produced and we place a response burden on them. Therefore, feeding data back to them for their needs is pivotal. To give three examples of the services that continues to bring value to companies and state institutions.

*To provide an example* the Labour Force Survey is conducted every year, the survey samples 85.000 people in the ages 15-74 years. A large variety of labour related variables are collected allowing for substantial analysis. However every year research institutions, private companies and public organisations finance the collection of extra variables – as a customized data solution against private financing. In these cases an extra module of data-collection is often created that allow them to answer their particular questions. For example how are disabled people doing in the workforce, what is their specific situation?

*Another example* is the anonymised microdata for state institutions that we offer from the registers that are relevant to the mandates of these institutions. The institutions are allowed to add their own register data to their projects – also in anonymised form. This allows ministries and other state institutions to conduct their own advanced analysis to support decision making. The projects are scalable depending on the institution’s needs – a flexibility that is brought about by the fact that they are customer-financed directly.

*Finally,* there are also hundreds of projects demanded that allow companies to access aggregated information relevant to their activityoften in smaller geographical areas,and the municipalities who access customized aggregated data that allow them to benchmark against other municipalities on their specific needs, for which the common feature is that they would not have been possible without direct financing.

The newly agreed PSI-directive from the European Union also addresses the subject. Income Generating Activities often, though not always, lie under the parachute of the activities in the PSI-directive. This directive addresses the reuse of public sector information – those that are made free as well as those that are delivered against a fee. The Impact Assessment of the PSI directive conducted prior to the revised directive measures significant impact of PSI reuse (where some are free and some aren’t) as

* Increased number and economic activity of Small and Medium sized Enterprises
* Improved API solutions (using data in applications)
* Up to 4% growth in the data market and growth overall including data related jobs
* Improved AI solutions

(Impact Assessment of the PSI-directive, Brussels 25.4.2018, Commission Staff Working Paper, European Union, Avialable at: <https://ec.europa.eu/digital-single-market/en>)

On top of implementing the act governing Statistics Denmark, Income Generating Activities in Denmark also form a key role in the provision of data as per the PSI-directive.

This very directine however, also states the key challenge to Income Generating Activities, namely the fee against which services are produced. The same Impact Assessment argues that a no-charge or reduced fee setup would increase economic actitivty, reduce entry barriers and limmit the uneven playing fiels created by the fact that some might afford what other might not.

This problem continues to pose a challenge to Statistics Denmark and the definition of which products to publish and which to deliver against a fee is a contiues discussion balancing among others other development-costs, user needs, enduse etc.

However within the existing framework, in total, we argue that these more than 750 customers and omore that 3000 individual services would not have been produced without the option of a fee.

*2.2. General methodological improvements and new developments*

*2.2.1. The case of new statistical products*

As we argue above these products are a gain in themselves, however we also argue that the fact that they are made feeds back into the official statistical production fundamentally strengthen it through *learning and methodological development*.

Some statistical products produced under the framework of Income Generating Activities, would most probably not have been produced under a publicly financed statistical programme.

While it might very well be possible to develop new statistics answering new trends and demands the exploration of new data sources, however, is often time consuming and new data sources are expensive to explore. Adding to this outputs are often uncertain and the outcome might very well be that a given statistical product cannot in fact be effectively produced anyway.

At the same time these new developments are often not a necessity for Statistics Denmark’s official statistical production and hence investments therein difficult to defend. They are *nice* to have – but not *need* to have. In this case users, whether public or private, can finance the exploration of new statistics against their tailor-made wishes and developments. These cases resemble a partnership-setup, since if the project is successful the end result is a new co-financed official and public statistical product. The exploration process, or pilot project, in itself is fully financed by the user and the final transformation into a public statistical product co-financed with varying levels of financing from Statistics Denmark.

As examples of Statistics Denmark engaging in such partnerships there are currently three ongoing contracts with the Ministry of Culture. The Ministry wants to increase access to statistics within their mandate. Specifically they want Statistics Denmark to explore statistics on book sales, the audio visual industry and live music. For each a pilot a project is setup with full user financing.

In the case of book sales, to provide an example, the catalyst is the access to scanner data from book shops, and the increased need for detailed information on sales for the book industry. While aggregate statistics have previously been produced on a yearly basis, an actual detailed statistic on title-level has not. Together with background data form the Danish Library Centre it is expected, that detailed statistics can be produced with data on sales channels, sales date, book format, genre, price, title and sales-location. The project entails exploration of data sources in very close cooperation with a significant number of private book industry actors.

Should the project result in the viability of a statistic on book sales, a co-financing solution will be sought allowing for the operation and publication of a new statistic.

The second case is of a similar nature. The catalyst here is the entire film industry’s need for data on the film industry economy, turnover and income sources – along the entire value chain. In this case a methodological exploration must be done in order to define the end-product and method for such new statistic. Financed by the Ministry of Culture an exploratory project is set up with the objective of developing a methodological paper that is the first step towards the production. The paper and end-product definition is developed in close cooperation with industry stakeholders, experts and researchers.

If the project is fruitful a co-financing solution will be sought that allow for the operation and publication of a new statistic on the economy of the audio-visual industry across the value chain.

In both cases, the provided option of customized data solutions against direct customer financing, feed into the development of new statistics within the official statistical program. The projects would not otherwise have been possible, within the budget of the institution – and the official statistical products resulting from the projects would not have been produced.

*2.2.2. The case of new methodological developments*

Just like in the case off book sales, Statistics Denmark is boosting the official statistical program with new development by engaging in a partnership to build a complex data warehouse on business statistics through the use of the privately financed customised statistics modality.

In this case the catalyst was researchers need for consistent time series, longitudinal register data and variables across a large amount of statistics and registers. A major project - funded 50 per cent by the Ministry of Research and 50 per cent by Statistics Denmark - is underway during the next five years to develop this data warehouse. The project is of a nature and scale that means that it would not have been possible to implement otherwise. In total the project is expected to cost around 2 mio. euros an investment that would not have been possible without external financing.

The project’s outcome is expected to create significant added value to researchers and customized statistics due to the improved microdata. However the data warehouse also represents a chance to break down statistical product silos between the involved official statistics. In the first instance the data warehouse involves the Statistical Business Register and the following surveys: ‘Production and Turnover in Manufacturing Industries’, ‘External Trade in Goods’, ‘Research, Development and Innovation’, ‘Enterprises’ use of IT’ and ‘Account Statistics (SBS)’. Obviously, the warehouse will enable integration with social statistics data.

This breakdown of silos, the increased integration of statistics and the use of data confrontation including at unit level is expected to affect the official production in itself in the long run. Hence the project is expected to provide a better understanding and picture of business units, a positive spin-off on improved sample coordination across statistics and a better basis for various analyses among others on globalisation.

In order to maintain the data-warehouse in the long run it is expected that certain methodological changes will need to be implemented at the level of production for the statistical program. Examples of such changes are expected to be changes in validation and imputation methods, due to the fact that both will be affected by the need for increased cross-consistency.

This project is a key element to reach a new milestone for business statistics, and it was driven and incentivized by researchers.

However just like this project is a milestone for the development of the business statistics, the question of cross-subsidizing the statistical production, has been raised. The question could naturally be raised, as to why this project is not financed as part of the statistical production.

However, due to the magnitude and the *nice-to-have* character of the project – it is argued that it would not have been possible to raise such budget (2 mio. EUR) from the official statistical production.

*2.2.3. The case of data confidentiality and its possible methodological improvements*

In customized statistics, crossing otherwise non-related variables, is often the objective. In this work - whether aggregated or disaggregated - a need for statistical disclosure control is created, which requires significant attention to detail and method.

In statistics Denmark Income Generating Activities has resulted in a new focus on data confidentiality and disclosure methods. Currently we are reviewing our policies and operational rules for disclosure control in order to provide the basis for delivering more detailed data to customers without disclosing data on individual persons or companies.

This work includes a monitoring of other data sources which could be used by customers as a supplement to the statistical data in a way that allows them to identify individuals.

The work with defining a set of minimum standards and an ‘operational guide’ for handling discretion issues has originated in the ambit of customized statistics, and is urgent due to the nature of the sales, however it is very helpful in the ordinary statistical data production as the challenges often are identical.

*2.2.4. The case of extra fault finding*

Due to budget cuts the last years Statistics Denmark has undergone and is undergoing a reform process on e.g. validation and editing named ‘Standardized and Modernized Validation and Editing’ which among many other things lifts the level of validation to what is necessary for the official aggregated level data production. The increased use of imputation and automatic corrections on less significant errors and missing data means that microdata are checked less than previously (but more systematically).

To ensure that microdata on key statistics are ready for tailor-made data production, whether for direct provision of microdata or aggregated tables, the price on income generating activities includes a portion for specific microdata validation on selected high-demand registers, known to be problematic if not subjected to extra checks.

This microdata check is not necessary to the official aggregate production, however the fact that the extra checks are made, positively influences the official statistical production and puts data to its maximum use.

Concluding on the issue of putting data to a maximum use and deriving maximum benefits from data one could definitely argue that this argument holds. The 754 customers served last year, with over 3000 individual services would not have been served. However minimising barriers to market and the avoiding an uneven playing field for enterprises must be a focus for NSI’s – where transparent and fair pricing is a strong tool.

Furthermore we have argued that the option of purchasing customized data solutions provides benefits to the organisation through new statistical products, methodological improvements and feedback. These products or projects, we have argued, are often risky in nature, they are not generally demanded and they are costly – which means that they would not likely have been produced without the option of private direct financing.

1. **Ensuring financial sustainability and transparency**

In 1966, when Income Generating Activities where introduced into Statistics Denmark’s mandate, one of the key arguments for Income Generating Activities was that these activities would be financially auto-sufficient and that one would save a series of costs, that would now be held by others. In other words, the financial aspect was also a factor.

In the below we analyse these financial aspects of the income generating activities closer.

The financial management on Income Generating Activities is governed by strict rules as to pricing and balancing, creating a need for strong financial management. A key element is that turnover and costs must balance to avoid adverse effects on the institutions economy but also to avoid cross-subsidizing the private demand with public funds aimed at the official statistical program.

The turnover stemming from privately financed customized data solutions has reached a level of approximately 100 million DKK, and now constitutes around 23 pct. of our yearly budget, amounting directly to approximately 89 man-years.

**Figure 1. Turnover over time**

So, looking strictly at the objective of financial sustainability and auto-sufficiency we can argue that the original assumption regarding Income Generating Activities stands and that the services do in fact add value to Statistics Denmark. However several challenges have also become increasingly apparent.

One of the most important challenges posed by the significant share of private financing in the institutions total budget versus government provision is the risks posed by *financial dependency* and possible *liability*. In this regard Statistic Denmark has in fact experienced situations where the collective finances have been affected when budgets have not been reached and expected income from overhead is not earnt. This in turn affects basic cost-sharing and the financial manoeuvring-space of the organisation and result in adverse effects on the institutions collective economy.

Moreover several years of unexpected deficits has obstructed the requirements for financial balancing and put financial management to the test.

Nevertheless, while balancing and managing the finances has produced several challenges, the organisation has been able to manage it and the organisation still has a significant turnover servicing those more than750 customers previously mentioned - therefore this risk is accepted and sought addressed through careful and conservative budgeting.

Asecond challenge that we face concerns the requirement to segregate income types for financial management purposes to avoid cross-subsidizing. In this regard the administrative rules around income generation activities versus those financed by the government appropriation are strict in order to ensure that cross-subsidisation does not occur.

Cross-subsidisation occurs either when customers’ resources indirectly or directly finance the official statistical production or the other way around. It is a risk that is battled continuously, but one battle that requires an overall understanding of the organisations activities.

To meet the financial risks faced Statistics Denmark has undergone a continuous administrative and financial centralization during the last 10 years allowing for a professionalization of the financial management, resulting in an increased transparency of financial management and increased professionalization of the administration.

Finally a new strategy for Income Generating Activities argues for further centralisation of financial management in the hope that this will improve outputs more.

1. **Other related effects**

Lastly, in an analysis of the benefits and challenges of Income Generating Activities as regards to a statistical organisation, an additional effect should be mentioned – one that concerns the added value of liaising with stakeholders that are non-traditional.

Many, if not all, official statistics domains have user groups in Statistics Denmark. Major ministries and institutions are invited. However, the user group that emerges by looking at the demand for customized data solutions is much broader than the ones we have in the user groups for the official statistical program and provide additional insight into the alternative end use of data. The Business data-warehouse is a good example of this.

*An additional example* is a major contract that runs on a yearly basis, under which Statistics Denmark provides several key variables on small geographical areas/clusters of around 150 households each. In this case the customer uses the data as input in a much larger production that involves a series of other data sources..

Due to the nature of their end-use, even small data breaches or methodological changes between years have major effect on their end models produced. The knowledge and understanding of these end users’ needs and challenges also allow for a feedback into the statistical production – ultimately improving it.

1. **Concluding**

Finally concluding on the original question that we have posed namely ‘*to what extent do the main benefits of privately financed Income Generating Activities outweigh the threats of the same and support these as a central element in the business of a National Statistical Organisation’,* we argue that the general benefits of the option of engaging in Income Generating Activities in an overall perspective significantly outweigh the threats.

Generally Income Generating Activities can improve the core business of statistical organizations. New methodological improvement, piloting and developing new statistical products and even better microdata are important benefits – that would not – to this level have been sought. Methodological feedback with better ways of dealing with data confidentiality issues and better dialogue with user are some of the other benefits.

Nevertheless, everything comes with a price. Bigger dependency financially, and a need to manage the risks associated with these activities, are perhaps the biggest challenge for NSI’s. We have argued that the financial requirements and the capacity to manage financial and administrative risks pose significant challenges to the organization. Nevertheless we have also argued that the risks are manageable, and that in a general perspective the benefits materializing in 754 customers served, more than 3000 services delivered that bring value to society – together with the methodological feedback mechanisms – bring more value to the statistical organization that they bring risks.

You could argue that new data solutions should emerge from central funding through the state budget to NSI’s. But in a time with budget cut-backs and general lack of funding, engaging in Income Generating Activities’s is a viable way to get the statistical gold mine of data of to work for policy, for research as well as for economic development.

Finally a set-up where the customers pay for the data solutions lead to more efficient ways of producing data – as customers require and demand transparency and efficiency throughout their meeting with the organization – a benefit and development that again strengthens the organization.

In conclusion it can be argued that the benefits of engaging in Income Generating Activities outweigh the challenges, however it can also be argued that in doing so careful attention must be placed on transparency of finances and administration, clear and regulated guidelines, careful and conservative budgeting and adverse market effects must be a continuous focus.

* 1. *New Strategy paves the way*

In this dilemma and to answer the dilemmas and challenges posed by the use of the Income Generating Activities in the NSI Statistics Denmark has developed a Strategy for Income Generating Activities for the period 2019-2022. The focus in this Strategy is not to increase the income Income Generating Activities, but mainly a response to the growing demand for data outside the organization and a response to the new possibilities the present technological era possesses.

The strategy aims at grasping the opportunities that the data landscape and data markets pose, while also meeting the threats and challenges that the organizations faces.

Moreover, the new Strategy tries to answer some of the problems we face when at least 25% of the budget comes from Income Generating Activities. It tries to answer these problems by identifying the demands, creating a vision, stating some principles and highlighting five strategical goals we needs to achieve order to keep up with the outside demands without compromising our identity as a NSI.

Statistics Denmark’s strategy for Income Generating Activities states that among other subjects the idea of Big Data and the growing demand for micro-data also from other than researchers is an opportunity that we need to grasp. It states that the principles for Income Generating Activities in our NSI is independence, credibility, transparency, data confidentiality, demand driven approach and the respect for existing markets and the behavior on monopoly markets.

This has lead us to the vision that Statistics Denmark’s Income Generating Activities is based on demanded and contemporary statistics and data solutions and skills, which can be seen as a natural extension of the official and public statistic production. By this, we will support the value creation in society by delivering facts and knowledge that substantiates the decisions, debate and value added business.

*To achieve this vision we have established five goals in the strategy*.

1. Our products are relevant and contemporary
2. Our processes are simple, effective and transparent
3. We have a clear communication of the possibilities and the products
4. Internally we cooperate to support quality in our products and in handling of our customers
5. We have a constructive cooperation with customers and we engage in partnerships both nationally and internationally

With these goals in mind we believe that we can face the future challenges and using the possibilities in the Income Generating Activities to support the core business of the statistical organization.

1. **References**

Impact Assessment of the PSI-directive, Brussels 25.4.2018, Commission Staff Working Paper, European Union, Avialable at: <https://ec.europa.eu/digital-single-market/en>)