Firstly, thanks to the NSB for the opportunity to participate in this seminar. Access to data and the integration of different data sources within a coherent framework is a subject of major importance to the Central Bank. In fact, we in the Bank are currently grappling with many of the issues raised, albeit on a much smaller scale to what is being discussed today. Over the years, individual data systems have been created by the different functional areas of the Bank– these systems were created independently of each other in different areas of the Bank. This fragmented approach has made linking different sources very difficult, both from a technical and methodological standpoint. To address these challenges, the Bank has developed a comprehensive data strategy where we are aiming to develop an integrated data architecture and governance framework. At least, as a single institution, the challenges, while daunting, are small in comparison to the development of a national data infrastructure. Nonetheless, this strategy will take a number of years, and a significant financial investment to implement.

We fully subscribe to the idea that data is a public good. Good quality data supporting extensive analysis is after all the life blood of decision making. One of the lessons of the crisis is the need for good quality, timely and reliable statistics. The recent Central Bank Macro-prudential rules provide a good example of policy which was developed, tested and supported by analysis and a detailed granular dataset. The Governor has publicly stated that any changes or recalibration of these rules will need to be based both on high quality analysis and a high evidence threshold.

The merging of data sources from different silos across the public service can provide much deeper insights into social and economic developments. In particular, the merging of information can hugely enrich policy debates. This,

however, requires collaboration between bodies on data requirements, standards, technical infrastructure, and common identifiers. In this context, I am pleased to say that the Bank and CSO have broad experience of these challenges based on a long record of working together

My comments today will focus largely on data available in the Bank and how this might contribute to enhancing the data infrastructure in Ireland. As well as our regular statistical releases, the Bank is developing a number of new granular data sources which will enhance information on the financing of household and corporate activity. Firstly, the Central Credit Register (CCR) is a major project aimed at supporting the Bank's financial stability, supervisory and consumer mandates. When operational, the CCR will act as an important support to lenders, showing an accurate picture of each borrower's total loans plus any guarantees provided. This matching of loans and guarantees within a Single Borrower View will facilitate enhanced creditworthiness assessments and responsible lending. Borrowers will also be able to check that information held in the register relating to them is correct. The Bank has engaged in extensive discussions with the Data Protection Commissioner in terms of implementing the necessary protection of personal information within the Register. While information derived from anonymised CCR data, will in itself be very valuable, the contribution to informing policy would be enhanced if this could be matched with other information, and particularly data held in CSO. The inclusion of the PPSN facilitates matching across data sets. If these data sets can be matched and anonymised, the potential for policy relevant analysis is significant. However, it must be stressed that any such initiative is subject to satisfying legal requirements on data protection and the approval of the DPC.

I will give a couple of brief examples of potential research, if data sharing is permissible. For instance, matching data on household finance from the Credit Register with data collected by the CSO under the Household Finance and Consumption Survey or with income or welfare statistics could significantly enhance information on different categories of households. This analysis can be done at regional level (e.g. by using the Eircode) or for particular target groups within the population. Secondly, the second phase of the Credit Register, commencing September 2018 will also include data from moneylenders. If matched with other data sources, this would allow analysis on the characteristics of households using money lending services. The great strength of granular information is that they allow distributional analysis not available from aggregate statistics. This was particularly obvious during the crisis, when aggregate statistics often masked severe difficulties for particular sub-sections of the population.

Obviously, the strongest safeguards must be in place to protect confidential information, whether relating to persons or to sensitive commercial data. However, the data needs of policy makers in particular areas may not require confidential information. Matching datasets and providing anonymised information will suffice in some cases – this would be particularly helpful for the Bank as it conducts research in a range of fields including economic analysis, consumer protection and financial stability. Appropriate structures and safeguards need to be established for merging and anonymising data from different sources – CSO with its long established tradition of handling confidential data and its access to administrative data should be the key player in bringing these different data sources together.

On the subject of confidentiality, I also think it is important to differentiate natural persons from legal entities. While data protection safeguards apply primarily to individual entities, a lot of information on legal entities or corporates is already in the public domain, through for instance the Companies Registration Office. I would strongly encourage some review of how confidentiality for legal entities is interpreted and whether current definitions are too restrictive. Obviously the highest levels of protection must apply to personal information or sensitive commercial information – however, should high level information on companies (such as name and address) be subject to the same level of confidentiality, especially if this information is available from other sources. The development of a register of companies, available to the public (not just public entities) would be hugely beneficial for a number of policy reasons and is an absolutely fundamental first step in developing a National Data Infrastructure. This brings me to the second major data project in the Bank which I will mention this morning. Anacredit (short for Analytical Credit datasets) involves the establishment of a comprehensive database and unique identifier for non-personal borrowers who have loans outstanding from Irish banks. This essentially equates to all legal entities resident in the State – in fact a register of corporate borrowers. The importance of comprehensive registers with unique identifiers is highlighted by the Nordic system which has registers at its core of its National Data Infrastructure. While we appreciate that legal restrictions exist in terms of accessing the CSO register, the Bank would strongly support any initiative which would facilitate the development of a public register of legal entities, which is an essential first step towards developing an NDI for Ireland. In fact, the current restrictive legal framework obliges the Bank to duplicate work already done in CSO.

Again, like the Credit Register, AnaCredit offers huge potential in terms of enhancing policy relevant information on business, if legislation allows merging with other data – for instance, analysis on employment or investment or on economic activity by sector, region, size of firm etc.

It may seem obvious, but for the development of the NDI, the importance of dialogue between users and data compilers cannot be understated. The challenge for compilers is to clearly explain what is available and how these data can be used. For users, the challenge is to define precisely what their needs are. The required outputs can only be delivered if the relevant source data are available. From experience, maximising the potential uses of data is enhanced if the following two criteria are met:

(1) data is collected on a granular basis as far as possible making it suitable for multiple user needs ; and
(2) data is collected under a legal framework that supports multiple uses and sharing, where appropriate.

In conclusion, the Bank strongly supports steps towards developing an NDI, subject of course to any legal constraints. We see data as a public good, essential for policy makers to monitor social, economic and other developments, and for private citizens to inform decision making and service delivery. It is not unfair to say that we are all facing demands for more detailed and timely information. Sharing data across entities as far as possible can help address these demands. However, the challenges in developing the necessary infrastructure and legal framework to support data sharing are formidable. The possible downsides arising from threats to data protection have been well publicised, while the positives in terms of enhanced information for policy making and a more informed society have not attracted similar levels of

attention. This seminar goes some way to balance the debate and the NSB can be assured of the Bank's support in its work to develop a coherent data infrastructure for Ireland.

Thank you for your attention.