

Christa Larsen, Sigrid Rand, Alfons Schmid, Mario Mezzanzanica, Silvia Dusi (Eds.):

Big Data and the Complexity of Labour Market Policies: New Approaches in Regional and Local Labour Market Monitoring for Reducing Skills Mismatches

Rainer Hampp Verlag, München u. Mering 2015, 307 S.,
ISBN 978-3-95710-045-0 (print), € 32.80
ISBN 978-3-95710-145-7 (e-book pdf), € 29.99

The experience of regional and local labour market observatories shows that information on some aspects of the labour market – such as the demand for skilled labour in certain sectors or spatial units – is difficult to obtain. In the recent years, ICT-related innovations have created new forms and types of data that can be used for enhancing the efficiency in several areas of economic activity. So far, the vast amount of unstructured data contained in the World Wide Web – Big Data – has been largely unexploited. However, as the available technology provides increasingly cost-effective solutions, it has become possible to provide services that have formerly been too expensive. Therefore, applying Big Data in labour market monitoring can provide innovative insights into the functioning of labour markets. Also the process data of Public Employment Services or Statistical Offices constitute a promising source of large amounts of data. The results of the analyses based on the different sources of data can be used to improve the efficiency of the labour market at large and the provision of services by governments and private enterprises.

However, the attempts to use Big Data in the context of labour market monitoring have been relatively rare so far, even though a growing interest can be observed among researchers and practitioners alike. Against this background, the issues of collecting, elaborating, analysing and disseminating the information available on the Web urgently needed to be addressed – as did the associated ethical and legal issues concerned with data ownership and protection. This year's Anthology of the EN RLMM covers these issues from the viewpoint of labour market researchers and practitioners in labour market observatories from different European regions and localities. The contributions provide first insights into new models and tools of labour market monitoring based on the usage of Big Data.

Key words: Big Data, Big Data application, regional and local labour markets, labour market monitoring, applied labour market research

Christa Larsen, Sigrid Rand, Alfons Schmid,
Mario Mezzanzanica, Silvia Dusi (Eds.)

Big Data and the Complexity of Labour Market Policies:

New Approaches in Regional and Local Labour
Market Monitoring for Reducing Skills Mismatches

Bibliographic information published by the Deutsche Nationalbibliothek

Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available in the
Internet at <http://dnb.d-nb.de>.

ISBN 978-3-95710-045-0 (print)
ISBN 978-3-95710-145-7 (e-book)
ISBN-A/DOI 10.978.395710/1167
First published in 2015

© 2015 Rainer Hampp Verlag München, Mering
Marktplatz 5 D – 86415 Mering, Germany
www.Hampp-Verlag.de

All rights preserved. No part of this publication may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publisher.

In case of complaints please contact Rainer Hampp Verlag.

FOREWORD

One of the important roles of the European Network on Regional Labour Market Monitoring is to identify and share best practices as well as innovative approaches and methods for monitoring regional and local labour markets across Europe. Cedefop follows developments in the EN RLMM with great interest as the Centre pursues similar goals through its work on the anticipation of skill needs, in accordance with the strategic objectives of the EU.

Analysing Big Data is one of the pioneer fields of research and practice that can be used to analyse the labour market trends and evaluate labour market policies. The main value-added of this field is that labour market information goes far beyond available official statistics. Thus, Big Data has the potential to provide a closer and immediate picture of skills supply and demand. Examining the potential of this type of data can provide further and much needed evidence to shape employment as well as education and training policies, including vocational education and training.

This is why I am pleased to write to short note to this year's EN RLMM Annual Anthology.

Joachim James Calleja

Director, Cedefop

CONTENTS

FOREWORD

Joachim James Calleja **5**

INTRODUCTION

Christa Larsen and Sigrid Rand **11**

1. BIG DATA AS A NEW DATA SOURCE FOR IMPROVED REGIONAL AND LOCAL LABOUR MARKET MONITORING

1.1. Big Data for Measuring Demand

Big Data Meets Web Job Vacancies: Trends, Challenges and Development
Directions

Silvia Dusi, Fabio Mercorio and Mario Mezzanzanica **31**

Monitoring of Internet Job Vacancies in the Czech Republic

Marta Salavová **45**

Big Data in Russian Sociology: First Steps Towards Understanding and
Co-operation

Svetlana Varlamova and Natalia Sedova **57**

Analysing of the Labour Demand and Supply Using Web Mining and Data
Mining in Romania

Claudiu Brîndaş and Ciprian Pânzaru **71**

Examples of Applying Big Data in Labour Market Analysis: The Małopolska Region

Katarzyna Antończak-Świder and Maria Leńczuk **87**

1.2. Big Data for Measuring Supply

Monthly Unemployment Rate Prediction with Google Trends Data:
Does Google Search Data Improve the Nowcast of the Italian Labour Market?
Monthly Unemployment Rate Prediction with Google Trends Data

Andrea Fasulo, Michele D'Alò and Stefano Falorsi **99**

The Value of Big Data in the Job Recruiting Process: New Opportunities
for the Italian Labour Market?

Renato Fontana, Vera D'Antonio, Martina Ferrucci and Carmine Piscopo **115**

The Use of Social Media in Human Resource Management: Evidences from
a Sample of Italian Companies

Mattia Martini **139**

The Use of Big Data: Challenges and Perspectives in Russia

Nina Oding **153**

2. BIG DATA APPROACH IN PRACTICE

2.1. Integrated Systems for Information Management in Regions

Planning Vocational Training for Employment in the Basque Country Using
a System of Indicators Combining Statistical Operations and Administrative
Registers with Information on the Labour Market

Javier Ramos Salazar **165**

Applying Big Data in Adult Education in Bosnia and Herzegovina

Željko Rajić, Siniša Veselinović and Zvezdana Jelić **175**

Development of Knowledge about the Labour Market in the Context of the Functioning of Big Data in Poland

Marta Sosnowska

193

2.2. Dissemination and User Groups

Interactive Visualisation of Bottleneck Vacancies: Introducing a New Tool for Administrative Data Using Berlin as an Example

Britta Lüdeke

209

Lessons for Regional and Local Skills Forecasting Arising from the Work of the European Network on Regional Labour Market Monitoring

Andrew Dean and Sigrid Rand

223

3. METHODOLOGICAL AND LEGAL CHALLENGES IN THE APPLICATION OF BIG DATA

3.1. Ownership and Individual Rights

Using Big Data from Multiple Sources in the Social Domain

Harry Piepers

235

3.2. Quality of Big Data

The PES Job Portal as a Labour Demand Indicator

Andreas Mångs

247

Assessing Data Collection and Data Quality for Labour Market Monitoring

Moreno Baruffini

255

4. INTERNATIONAL PERSPECTIVE ON THE APPLICATION OF BIG DATA IN REGIONAL AND LOCAL LABOUR MARKET MONITORING

The United Nations and Big Data in the Strategy for Global Sustainable Development

Marco Ricceri **275**

INFORMATION ON AUTHORS **291**

INTRODUCTION

Christa Larsen and Sigrid Rand

For many years, the European Network on Regional Labour Market Monitoring (EN RLMM) has been exploring how regional and local labour market observatories can provide optimal support for decision-makers. Creating transparency on the existence of specific data, supporting their interpretation and deriving corresponding action strategies are meanwhile indispensable building blocks of a successful monitoring approach. However, establishing transparency in regional and local labour markets is often a demanding task, because on this small scale less data are available than on the national level. At the same time, there are only limited resources available for generating new data stocks. Therefore, considerable data gaps arise. This is especially critical, because decision-makers on the regional and local level need exact and detailed information.

Particularly in the last five to eight years, the members of the EN RLMM have made considerable efforts to explore how these data gaps in regions and localities can be overcome. In doing so, a special focus has been on the scarce financial and personnel resources of the observatories. Two approaches for generating further data have proven efficient and effective regarding the explanatory power of the information (see the EN RLMM anthologies 2007-2014¹):

- Using diverse process and social security data as well as data collected by educational organisations and administrations,
- Including expert knowledge in the monitoring process.

Process data are generated in a multitude of administrative processes of the Public Employment Services (PES) (e.g. data on unemployment, activation and qualification measures as well as the duration of unemployment and the vacancies). Further data sources include social security data (e.g. employees subject to social security contributions) as well as data generated by educational organisations and administrations (e.g. number of graduates, number of people in education), which can be broken down to small-scale level. Since these are population data, limited representativity does not pose a challenge. However, the high level of data aggregation and summing up of categories to comply with data protection rules constitute the main challenge of working with this data stock.

¹ The list of EN RLMM publications is available at: www.regionallabourmarketmonitoring.net/publications.htm.

Consequently, these data do not have the level of detail required by the decision-makers. Against this background, (mostly) local experts are used as a complementary source of information. Therefore, the involvement of regional and local expertise is an important component in the monitoring approach represented by the EN RLMM. However, the use of expert knowledge always entails methodological challenges, especially considering the objectivity of this singular information.

With the help of these approaches – process data and expert knowledge –, data situation in regional and local labour market observatories in Europe has been furthered quantitatively and qualitatively. Despite these positive developments, considerable data gaps remain. Consequently, local decision-makers are forced to rely on an “uncertain” data and information basis. Therefore, the members of the EN RLMM are constantly making their best endeavours to open up new data sources for regional and local labour market monitoring. Meanwhile, Big Data has become an important issue in politics and the economy. Researchers as well as representatives of the different fields of practice where data is created are increasingly addressing this topic. It is assumed that there is an immense potential in these data, so that the question arises, in how far Big Data can be used for regional and local labour market monitoring. The experiences that the Network Members have already made with Big Data and the associated chances and challenges will be discussed in this anthology.

1. What Is Big Data and Where Is It Used in Labour Market Research?

The term Big Data refers to data, which are generated through different digital devices such as smart phones, websites, apps, sensors embedded into objects, scanning of machine-readable objects (barcodes) and Social Media postings. One of their most important characteristics is, that they are accessible through the Internet (Kitchin 2014). These data can be used to depict behavioural, consumption and mobility patterns of individuals. The generated data pool grows with high velocity and has so far its potential for research and consultancy has hardly been harnessed. The concept of Big Data addresses not only its quantity as the term might suggest. Rather, Big Data is characterised by the multitude of data sources and high velocity of data analysis. These central characteristics are described as the three Vs (volume, variety, velocity) (Burns 2015, Mayer-Schön-

berger 2015). It refers to a situation in which a fine-grained picture or assessment of a situation can be reached through analysing large quantities of data from different sources and of varied structural characteristics (structured, semi-structured, unstructured) (Kitchin 2013, 2014, Mayer-Schönberger 2015). Only fairly recently, extracting and analysing these data stocks has become possible through the development of information technologies such as crawling, dictionaries and semantic tools. Through algorithms, correlations in the data are identified, which refer to patterns. This enables to describe patterns in real time as a kind of “snapshot”. (Kitchin 2014). However, in this process neither the influence of the context nor the historical development are captured (Burns 2015, Kitchin 2013). As a result, this snapshot is ahistorical and de-contextualised. Especially in social sciences and human geography, critical voices can be heard pointing out that these data cannot speak for themselves, i.e. they are not self-explanatory (Barnes 2013). Along this line of argument, it is assumed that these data only acquire the meaning – i.e. are suited to not just describing social phenomena, but also to explaining them – when they are placed in relation to theories, models, concepts or hypotheses (González-Bailón 2013, Kitchin 2014, boyd and Crawford 2012)². In order to demonstrate this necessity, a further feature of Big Data is introduced – value and veracity³. This should make explicit that the data as well as the results of analysis still need to be verified. Along these lines, also a 4Vs concept of Big Data is postulated (Reussner 2014, Kreibich 2015⁴).

Whilst the technical possibilities have largely determined the framework for exploiting Big Data so far, often based on concepts of measurement typical of natural sciences, social sciences are increasingly contributing to the discourses. On the one hand, they see the potential of Big Data; on the other hand, they make clear that further methodological, legal and above all ethical questions need to be clarified. As a methodological challenge, the representativity of Big Data needs to be considered – as well as the generalisation of the statements based on it. Meanwhile, considerable efforts are made to grasp the representativity of Big Data. Often the reference data from surveys or official statistics are used for comparison purposes, the degree of the bias is assessed and, if necessary, Big

² Robert Kitchin (2014) suggests feeding Big Data into the process of abduction, indication and deduction. This would enable to use its potential for categorising reality. At the same time, established research paths could be used to explain its relevance and meaning.

³ Dusi, Mercurio and Mezzanzanica in this volume use here the term “believability”.

⁴ Rolf Kreibich (2015) equates the 4 Vs model with Smart Data: Smart Data = Big Data + Utility + Semantics + Data Quality + Security + Data Protection). He denotes that this will lead to useful and scrutinised data of high quality.

Data is weighted. Similarly, expert interviews are used for comparing structural patterns. Further statistically informed techniques such as the analysis of the missing data are being developed (Kurekova et al. 2014). A more qualitative approach for capturing the quality of Big Data is directed towards determining quality criteria, which arise from traditional empirical social research such as timeliness, relevance and accuracy (see the contribution by Moreno Baruffini in this volume). Legal and ethical discourses address above all the issues of data ownership and hence their utilisation through third parties (Stöcker 2015). From the ethical perspective, the aspect of utilising individual digital trace data in the Social Media becomes relevant – especially considering that often data are used for purposes of which the individuals were unaware of at the time (e.g. recruiting). Resulting from the different legal practices in single countries, many users of Big Data currently work in a grey zone. As national borders constitute no barriers for Big Data, a common European law – the data protection regulation – for dealing with ownership is presently being developed (Stöcker 2015, Mayer-Schönberger 2015).

The above-mentioned complex challenges concerning the use of Big Data underline that its utilisation is an interdisciplinary undertaking requiring the participation of IT specialists, social scientists, statisticians, legal experts and possibly philosophers (González-Bailón 2013). In addition, it is brought to our attention that different techniques and tools need to be applied (e.g. multi-method approach). This becomes clear when we look at the process of acquiring and analysing Big Data. The selection of data sources (mostly from the Internet) needs above all expertise in the field of social sciences and law. In contrast, the extraction of data with crawlers or robots is strongly IT-driven. However, already the subsequent process of data management, i.e. data cleansing and data preparation requires besides technical input also the contribution of social sciences, for example when the data need to be sorted according to semantic categories or dictionaries need to be developed. The following analysis requires context knowledge in order to recognise patterns and interpret them. Here, approaches from social sciences and statistics become important such as the representativity check or relating it to theories and models. With a view to the use and dissemination of data or information, this means that the issues of data protection need to be clarified. The necessary interdisciplinary approach makes co-operation between the representatives of different disciplines necessary – also across the differences between the disciplinary traditions and cultures.

Despite these complex challenges related the extraction and use of Big Data, its utilisation is promoted for several reasons:

- It is clear to everybody that the digitalisation of our everyday life is not a reversible process. It is important to observe this development and adjust to it. Especially in the social sciences it is argued that these developments need to be studied and shaped (Graham and Shelton 2013). In doing so, ethical and social aspects concerning individuals are placed in the centre of interest (boyd and Crawford 2012). At the same time, it is obvious that digitalisation can have fundamental effects on social relationships and forms of communication as well as social developments (see, for example, Fontana et al. in this anthology as well as Gonzales-Baiton 2013).
- From the viewpoint of users, the central advantage of Big Data is that big amounts of real-time data can continuously be acquired, thus making this process inexpensive. Consequently, even if public budgets are experiencing significant cutbacks, research and consultancy will still possible.
- Big Data enables to depict and analyse the behaviour of people in the context of their life. This enables to avoid biases, which so far have brought about undesired effects in empirical designs, e.g. through social desirability.
- Through Big Data – especially through the fine-grained situation analyses – decision-makers can be optimally supported in their decision-making process (Gonzales-Baiton 2013).

On the other hand, there is the view that the drawbacks of Big Data including limited representativity, ethical and legal questions are still to be addressed (Welles 2014).

So far, Big Data has been above all deployed in the field of banks, insurances, health, energy, industry, transport and IT and only a few approaches can be found in the (regional) labour market monitoring (Askatas and Zimmermann 2015, Kreibich 2014). The Prediction of Unemployment analysis is farthest developed. This is the result of the need for short-term forecasts of unemployment in the aftermath of the global economic crisis from 2008 onwards. Technically, these analyses are based on correlations, which are identified in combination of particular key words in search engines and monthly unemployment rates. The observed structure is used for forecasting unemployment rates. This approach renders not only forecasts possible, but also the so-called nowcasts. This is an up-to-date differentiated situation analysis which, for example, shows which topics are relevant for the users of the Internet in times of high unemployment.