

Process generated Research Data from the Basic Income Support System

Insights into the IAB Development Process

Martina Oertel, Ulrich Thomsen (Institute for Employment Research)¹

The German Basic Income Support System in its current configuration was set up in 2005 by the introduction of the Social Code Vol. II (SGB II). The administrative implementation of the SGB II led to new institutional settings. Three different institutional types of agencies were created to administer the new system. On the one hand cooperation between Federal Employment Agency (BA) and municipalities can be found, on the other independent implementation by the municipality. The two administrative settings resulted in differences in the IT-systems supporting the administrative processes. The conjointly (called ARGE)² or cooperatively (called gAw) administered institutions after a transition period of a couple of month use centralized operative IT-systems³ (called A2LL), the independent municipalities that opted out of the cooperative administration (called zkT) can use any supporting software they prefer. The data from the zkT are transmitted to the BA via a standardized interface (called XSozial-BA-SGB II⁴) whose configuration is defined by BA in an open discussion process between the BA and the zkT including the central municipal associations.

¹ The authors hereby present the results of the work of several teams, because developing research data always needs a lot of specific knowledge and creativity:

LHG-Team in IAB department 'IT Services and Information Management': A. Athmani, B. Eichler, C. Erler, A. Jasim, M. Köhler, R. Konstanty, M. Oertel, U. Thomsen, H. Siami
Additional members of *IAB-workshop for SGB II – Data*: S. Bernhard, J. Dornette, T. Graf, E. Jozwiak, D. Oberschachtsieck, S. Popp, H. Rudolph, B. Schels, M. Trappmann, J. Wolff and many other temporary active colleges.

The LHG-Team has to thank BA-colleges (Statistic/IT-Systemhouse) for many advices: R. Bergdolt, S. Breuer, B. Hofmann, F. Gellern, K. Gillarek, J. Korn, T. Kuhnt

² ARGE: **A**rbeits**G**emeinschaft nach § 44b SGB II / gAw: **g**etrennte **A**ufgaben**w**ahrnehmung / zkt: **z**ugelassener **k**ommunaler **T**räger nach § 6a SGB II

³ gAw-agencies should provide information about housing costs via the Xsozial-BA-SGB2 interface.

⁴ Information about job-search and active labor market policy program participation is also provided via XSozial-BA-SGB II-Standard, but due to the concentration in benefit recipience is not considered further.

The “Benefit History of Means-Tested Unemployment Assistance” (LHG) as the central IAB research data set used up to now data from the A2LL IT-system but not data from XSozial. The first version of the LHG was developed until November 2006. After finishing, this first version was utilized by selected users and thereby intensively tested, resulting this summer in a review and afterwards a redesign of the data product. It is important to notice, that up to today no data from the zKT are integrated. Data from these municipalities are missing today and in the current IEB-Version so that there are regional and other limitations in that data.

These limitations will partly end next year, because the next version of the LHG in 2009 will contain all registered beneficiaries.⁵ Included in the next LHG are first individual variables and second structural information of the needy household (BG). In the following an overview over selected and important information in the LHG is given:

- Start/end of benefit recipience as a BG-member*⁶, exit reason
- Start/end of sanctions*
- SGB II – agency* and institutional type of agency, municipality of place of residency*
- number of BG-members*, total members of household*, type of BG*, individual role in BG*
- individual earning capacity*, different BG-members earning capacities*
- sex, birth date, nationality*, marital status*
- additional identifier and other technical information

What are not included are payment measurements. In former versions information about average payment levels could be found, but this information was cancelled after the review process. The reason for that lies in the fact that spells in the LHG are relatively long, whereas payment information often changes monthly. An average treatment of payment information over the complete spell duration blurs these differences so that a clear interpretation isn't possible.

The same problem holds true for income information etc. so that a different solution is chosen. Supplementary files for payment, income und sanction information are created. These files provide detailed information to the according fact and can easily be merged to the LHG-file. The drawback of this detailed information lies in its complexity.

⁵ Persons that according to §7 SGB II are registered and administered in the SGB II-system

⁶ A star (*) marks information that, after changing, generates a new spell

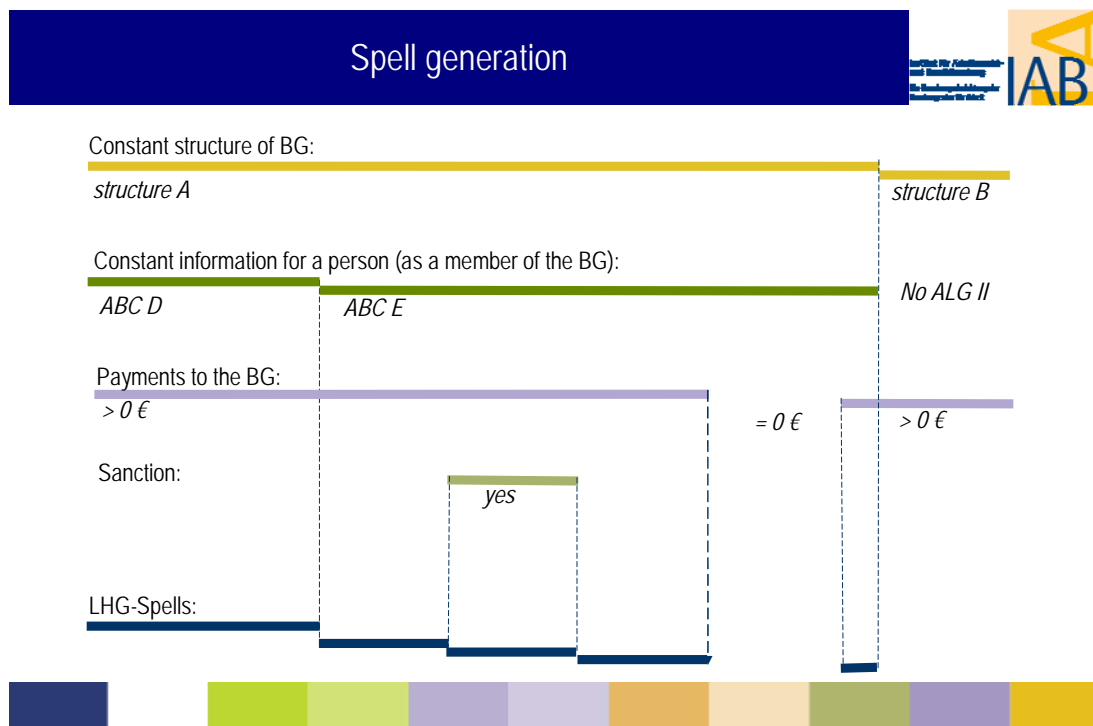
- Income for example includes 14 categories of different income, like employment, self employment, pension payments, child allowance, etc.
- The payment supplementary file includes detailed payment information about monthly payment in 30 categories, like standard benefit payment, different social insurance contribution, total housing costs, etc.
- In the detailed sanction data information like reason for the sanction, sanction time level of sanction and much more can be found.
- And last but not least, detailed housing costs are available in a supplementary file. Housing costs break the total costs down to e.g. level of rental fee, or incidental expenses.

When linking the supplementary files to the base file, the result will be a much larger data set containing much more spells per person as the base LHG. The reason for that lies in the fact that LHG spell durations are relatively long. In the following section the generation of spells respectively the splits is sketched out.

Compared to the existing IAB research data the introduction of the SGB II posed new challenges to the creation of research data sets. Before the SGB II the spell generation in the research data was focused on an individual person. Changes in the personal attributes, status information etc. resulted in a new time period in the data. With the introduction of the SGB II this person centered view was broadened. In the Basic Income Support System the Needy Household is in the central focus. This means to incorporate BG-relevant information into the older person centered logic. A change in the BG-structure may affect individual information like payment level or may even cause an exit out of the system. The important criterion for being in the system is being a member of a BG where at least one person receives benefits.

In the introduction of available LHG-information split generating attributes are marked by a star. The following graphic illustrates this logic:

Figure 1



One reason for the necessity of intensive testing and the review of the LHG is data quality. As mentioned above there are two sources for benefit recipient data in the SGB II-system. First there is the operational and central IT-system A2LL, second there are the data of the zkT (and partly the gAw), which are transmitted to the BA via a standardized interface⁷.

The different operational IT-systems (A2LL and several⁸ at the zkTs) cause different types of errors and problems in the data. The most important we want to point out here.

There might be gaps in the biography of a person because of missing notifications. This might be the case in data from all different types of agencies. The problem is extensive in 2005, less in 2006 and later on may be found punctually.

The problems in 2005 result from starting phase of the SGB II in the administrative processes as well as in the IT-system. Workarounds were common in the first month of 2005 and all agencies built up their administrative system step by step. That means that in the first months some SGB II benefit recipients from the zkT still were administered in the BA IT-systems and changed later into the domain of the zkT. But not only the zkTs used "foreign" systems in the first phase all other types of agencies did so too. It was not until spring of 2007 that all ARGES and gAws used A2LL

⁷ A detailed description of the interface as well as many more important information (XML-Scheme, different work tools) can be found under the following link: <http://www.pub.arbeitsamt.de/hst/services/statistik/000200/html/sgb2/index.shtml>

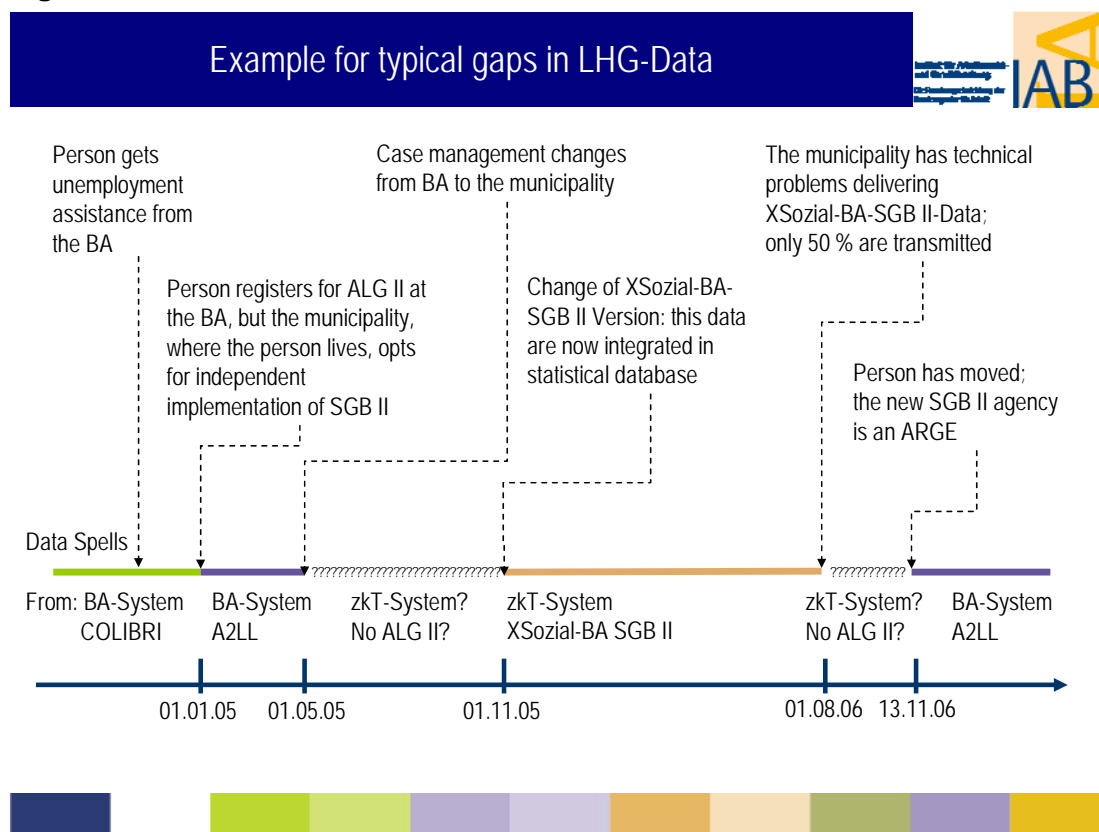
⁸ There are five software companies partly offering more than one product.

as the administrative tool for the registration and handling of the SGB II benefit recipients.⁹

On the field of XSozial-data there are two phases. Before November 2005 most of XSozial-data are not usable for building up a research data base due to inconsistencies or gaps in the data. They were used only on an aggregated level for producing statistical information. Between November 2005 and December 2006 the data got better, but were still not complete when collected and integrated in the BA.¹⁰ Until then the statistical data base in the BA is more or less stable and can be used to produce firm research data products. It has to be kept in mind though that still transmission problems exist and might cause missing data in some of XSozial-modules, therefore resulting in gaps in parts of the data in certain months.¹¹

All the described problems in the data might cause gaps in the biography of a person. This illustration makes some problems visible:

Figure 2



⁹ The Statistic of BA displayed the information of 255 counties as being complete in January 2005 meaning that nearly two thirds of the projected cases are complete (cf. Statistik der Bundesagentur für Arbeit (2005)).

¹⁰ Cf. Statistik der Bundesagentur für Arbeit (2007), p. 5 et sqq.

¹¹ Cf. Statistik der Bundesagentur für Arbeit (2008), p. 25: In the mean in 2007 of the total 439 counties (until June 2007) respectively 429 counties (since July 2007) the data of 1.3 counties were missing.

Beside the general availability of data of different data sources over time, there exists another problem in the data on individual level. Lacking or over time changing identifiers in the XSozial-Data result in complex algorithms for assigning information to a person as being one person. In theory¹² the major and comprehensive identifier of a person in A2LL and in XSozial should be the BA-client number. But there is no systematical possibility for the agents in the zkT to check if a new client already has a BA-client number. So if a client doesn't provide the BA-client number in the first registration process, there is only a little chance to find that identifier for the registering agent. The identification of a person as being the same person in older and newer XSozial transmissions, XSozial-data and BA-data has to be done in the BA by complex matching procedures. The identification algorithms have been changed several times over the last years, but now a good working algorithm has been found. Although, it cannot be excluded that in a small number of cases some information are incomplete or are a wild mixture of spells of different persons.

In the A2LL data the situation considering the identification of a person is much better. But the IT-system itself is not very user friendly and stable.¹³ Beside these two facts, there exist a number of temporary or lasting workarounds for a couple of problems or missing features this IT-system has. Data or data artifacts produced by such workarounds are partly hard to identify in the research data and therefore can be found only in parts in the LHG. Furthermore there are no complete reliable information in A2LL-Data about sanctions before 2007 and income between October 2005 and December 2006. From the beginning up to today workarounds were and are in use, whose results may not be identified and differentiated comprehensively in the data.

To close this part about the data quality it can be said that there are temporary or lasting limitations in the data not yet completely analyzed and documented.

To sum up the presentation of strengths and weaknesses in the IAB research data three questions will be answered:

– What's possible?

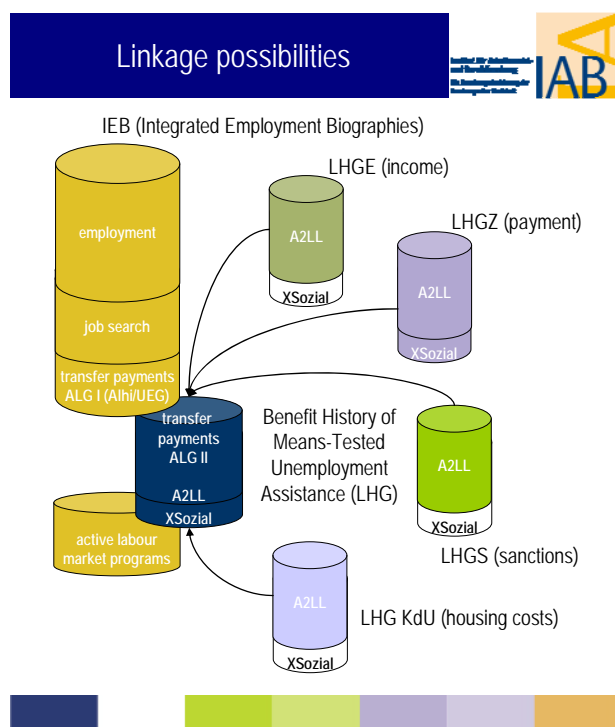
Flexible and changing views via varied linkages on the SGB II-benefit recipients, their situation, their BG and associated BG-members can be produced and analyzed.¹⁴ The following illustration shows the combination of available data sets:

¹² And defined by law: cf. § 51a SGB II

¹³ Cf. ISG (2007), p. 79

¹⁴ Examples for research results that are based on LHG-Data amongst others can be found in: Bernhard et al (2008), Bernhard and Wolff (2008), Hartig et al (2008), Hohmeyer et al (2006), Wolff and Nivorozhkin (2008)

Figure 3



As covering the complete target population it is possible to draw samples of rare combinations of attributes in an adequate size.¹⁵

– What's critical?

First, computing of durations and entrance/exit rates when 2005/2006 has to be included is critical because of the gaps in the data.¹⁶ The same also applies for identification of status and status transition when all the SGB II- benefit recipient data are integrated into IEB.

Second, partly higher rates of implausibility may occur in the SGB II-data (especially XSozial-data and again the period of 2005/06) compared to other data like employment. Thereby the work for data transformation for the researcher is extended.

– What's planned?

XSozial-Data will complete the LHG and LHGZ in first quarter of 2009. In second quarter we plan to integrate this data in IEB. During review processes we want to include XSozial-Data into LHGS/LHGE/LHG KdU until the end of 2009.

Important for external researcher is that a concept for a SUF hasn't been developed yet. This could be done after gaining some experience with the integrated data from both sources. Beside the experience in working with the data there is a new chal-

lence in anonymizing the data because of the new focus on BG-structure, information etc. So if there is a demand for the production of a SUF this will take some more time.

Literature:

Statistik der Bundesagentur für Arbeit (2005): Eckwerte der Grundsicherung für Arbeitsuchende nach SGB II nach Kreisen (Eckwerte) - Daten nach einer Wartezeit von 3 Monaten, Januar 2005;

<http://www.pub.arbeitsagentur.de/hst/services/statistik/detail/l.html>

Statistik der Bundesagentur für Arbeit (2007): Zwei Jahre Datenlieferungen der zugelassenen kommunalen Träger nach § 51b SGB II über XSozial-BA-SGB II. Entwicklung in den Jahren 2005 / 2006. Nürnberg, Januar 2007

Statistik der Bundesagentur für Arbeit (2008): Qualitätsbericht: Statistik der Grundsicherung für Arbeitsuchende nach dem SGB II. Nürnberg 16.05.2008

Bernhard, S. / Gartner, H. / Stephan, G. (2008): Wage subsidies for needy job-seekers and their effect on individual labour market outcomes after the German reforms. IAB Discussion Paper, 21/2008, Nürnberg

Bernhard, S. / Wolff, J. (2008): Contracting out placement services in Germany * is assignment to private providers effective for needy job-seekers? (IAB Discussion Paper, 05/2008), Nürnberg

Graf, T./ Rudolph, R. (2006): Bedarfsgemeinschaften im SGB II 2005: Beachtliche Dynamik bei steigenden Empfängerzahlen. IAB-Kurzbericht 23/2006, Nürnberg 2006

Hartig, M. / Jozwiak, E. / Wolff, J. (2008): Trainingsmaßnahmen: Für welche unter 25-jährigen Arbeitslosengeld-II-Empfänger erhöhen sie die Beschäftigungschancen? IAB-Forschungsbericht, 06/2008, Nürnberg

Hohmeyer, K. / Schöll, C. / Wolff, J. (2006): Arbeitsgelegenheiten in der Entgeltvariante. Viele Zielgruppen werden noch vernachlässigt. IAB-Forschungsbericht 22/2006, Nürnberg.

Hohmeyer, K. / Wolff, J. (2007): A fistful of Euros: Does One-Euro-Job participation lead means-tested benefit recipients into regular jobs and out of unemployment benefit II receipt? IAB-Discussion Paper 32/2007, Nürnberg

ISG (2007): Evaluation der Experimentierklausel nach §6c-SGB II – Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung „zugelassene kommunale Träger“ und „Arbeitsgemeinschaften“. Jahresbericht 2007, Köln 2007

Wolff, J. / Nivorozhkin, A. (2008): Start me up: The effectiveness of a self-employment programme for needy unemployed people in Germany. IAB Discussion Paper, 20/2008, Nürnberg

¹⁵ The LHG served as pool for sampling the survey of the §6c-Evaluation (ZEW / IAQ / TNS Emnid (2007) and serves now for the IAB-Panel PASS

¹⁶ An appropriate regional selection and possibly projections can help to deal with these problems (e.g. Graf and Rudolph (2006))

ZEW / IAQ / TNS Emnid (2007): Evaluation der Experimentierklausel nach §6c SGB II – Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung „Zugelassener kommunaler Träger“ und „Arbeitsgemeinschaft“; Untersuchungsfeld 3: „Wirkungs- und Effizienzanalyse“ – Erster Bericht. Mannheim, Gelsenkirchen und Bielefeld Juni 2007