

Addition of data to portals - selection, acquisition, processing and uploading of data

Introduction

The aim of the manual is to share experiences that INEKO gained during the creation of the portal about primary and secondary schools in the Slovak Republic (<http://skoly.ineko.sk>) and the portal about management of local governments (<http://obce.ineko.sk>) – specifically the experience relating to selection of, acquisition of, processing and uploading of data to these portals. Note: a newer version of these portals (in the English language) can be found at <http://schools.ineko.sk/> and <http://municipalities.ineko.sk/>.

At first, INEKO launched its portal about primary and secondary schools in late January 2012. The inspiration for its creation was the fact that the public had no comprehensive information about individual schools in Slovakia until then. For example, parents lacked help when choosing a school for their child, they had no easy way to get information about the results of the chosen school, and the teachers had no feedback on the results of their work.

With the aim of providing as comprehensive picture of individual schools as possible, we decided to prepare a portal that would enable the searching and comparison of schools according to various criteria, in a way of ensuring the highest possible user friendliness. Ideally, this portal should cover all schools in Slovakia (all school types), provide basic information (type and basic characteristics, contact information), but mainly qualitative pointers, in which case the developers can be the most limited by availability of data.

To provide an example, in the optimal case we would like to present information about the school's added value, i.e. about how much the knowledge of graduates a selected school improves in comparison to the knowledge they had when entering the given school. Such data are not yet available in the Slovak Republic; the aim is therefore to provide the largest quantity of the best possible data (considering their accessibility) that can in some way testify to the quality and conditions of education at individual schools, and visualize and present these data as simply as possible to the users of the portal.

With a similar objective, we launched a second portal (in May of 2012). In that case, the entities of interest were not schools, but local governments. The public thus acquired an instrument for monitoring the management of local governments and, thanks to that, citizens were given the opportunity to check how the financial health of their municipalities changed, for example, since the last municipal elections. Representatives of local governments were given the opportunity to compare their operation with other local governments of similar geography or size.

The procedure that we used in the search, selection, acquisition, and processing of data and uploading to the portals, will be discussed further in the following chapters. We do not expect that this procedure to be universally

applicable in the conditions of other countries, but the valuable experience we have acquired when preparing Slovak portals can be an asset and inspiration in the data preparation of portals in foreign countries.

The school portal, which is a little more complex, will be primarily used as an explanation of the entire procedure - we will use it for a detailed analysis of the procedure. Subsequently, in relation to the local governments' management portal, we will only list the differences in the procedure (compared to the procedure for the school portal).

School portal - selection of data

We based it on the assumption that the portal will be most useful to the public if it permits the finding of all data on the individual subjects (schools, municipalities) that could be useful for users, and in one place. In the first phase we have therefore made a summary of data that might be of interest to users - so far regardless of the real availability of data.

In the case of the school portal, we identified the following areas of information of interest (for individual schools):

- Contact information (name and address of school including telephone number, e-mail, web site address and GPS coordinates for later visualization of the school's location on a map)
- Type of school, for example. by age group of students (e.g. primary school, secondary school, gymnasium, conservatory of music)
- Language of instruction (including information on whether the school has a bilingual focus)
- Information on the administrator of the school (e.g. if the administrator is the state, private individuals or the church)
- Information on whether the school specializes in teaching physically or mentally disabled students (together with the specification of what education is provided for these students by the school)
- Information on whether the school has a dormitory
- Additional information about the school (the current number of daily and external students, the number of graduates over the past year, the number of students from socially disadvantaged backgrounds, the number of teachers, etc.)
- Conditions for teaching (human resources), for example: teachers' qualifications (certificates), language skills of teachers, the number of external teachers on work experience
- Conditions for teaching (technical), for example: schools equipped with computers, projectors, electronic whiteboards, internet connection speed
- Other school facilities (library, dining room, buffet, sport grounds, other opportunities for spending time after the end of classes)
- The financial resources of schools, divided into public funds, private funds (from parents), funds from awards for school achievements, funds from commercial activities (e.g. rental of premises)
- The results of inspections on schools
- The test results of students (for all grades and subjects where testing took place), and ideally in a way to enable comparing the results of individual students over time (subject to the protection of personal data)
- Student achievements in competitions
- Involvement of schools in international projects (e.g. exchange of students, teachers)
- Results of school graduates in the labour market (unemployment, average salary)
- Success of school graduates in the admission for further study (including an assessment of the quality of schools where the students were admitted)
- Information about the charges associated with study at the school

We tried to get the data for all schools and, where possible, for at least the five most recent years available.

Note: It is important that the data are not collected from the individual schools, but acquired centrally from an institution working in the field of education - the reason is to ensure the methodology does not vary for individual schools. If the data is collected from one institution for all schools, it ensured the uniformity of methodologies and comparability of data between schools.

School portal - data acquisition

We then compared our "interest" list referred to in the previous section with the actual availability of data on the websites of institutions active in the field of education in the Slovak Republic.

Where possible, we obtained the data directly from the Internet. The data that we failed to get from the Internet we then tried to obtain from the competent institutions using "info requests".

Note: "Info requests" are a type of request that are supported by Slovak law and are aimed at obtaining information (institutions belonging under the State should provide information, if it's not secret or e.g. in violation of the Act on personal data protection; but may not necessarily provide data, which they would have to process into the desired format, if they do not have it in this format).

Note: The real availability of data (whether due to their public availability, or due to their acquisition) naturally decide what set of pointers can be provided to users on the portal. The list of remaining pointers or data that could not be obtained and would be useful for the portal or the public, will not go to waste. It may serve as a basis for formulating recommendations to improve data collection in the field of education in the country. Such a list, along with the justification of usefulness of the proposed collection of additional indicators, may be sent as an inspiration, for example, to the Ministry of Education in the given country.

It can be estimated that we could not get more than half of the data we were interested in. In particular, we could not get the most significant data - the data on the basis of which it would be possible to assess the added value of individual schools. Specifically for this purpose, it would be necessary to acquire the testing results of students for various years. However, they are currently available only for the last year of primary and secondary schools. Thus, we do not know the level of knowledge of students at the time of admission to the school, but only at graduation, so we cannot assess the difference, the estimate of acquired knowledge. In view of this, we are able to capture the quality of schools only partially, and therefore we can only speak about monitoring and assessment of the results of their students.

It should be added that when acquiring the data, their form is also important. We were trying to get the data in electronic form, ideally in the form of databases - so that data can be processed as simply as possible. It is essential that each school can be identified in the best manner possible - this can be ensured either by stating the precise name and address of a school or by providing an identifier (e.g. a unique code number for each school in Slovakia), or by both at the same time.

School portal - "database of schools on the portal" and the pairing of databases

It would be ideal if all the institutions in the SR were using a uniform set of identifiers for schools and if the identification of schools using these identifiers was included with all databases. But the reality is that there is no single identifier for schools in Slovakia used by all organizations involved. Some institutions under the Ministry of Education use the code KODSKO, but not even the Ministry itself records all the data for schools along with this code. Some other institutions use their own codes for schools, and some provide the data without any codes at all, only with the names and addresses of schools.

This fact is a significant complication in the pairing of databases, because if we want to compile profiles of schools that will include all the interesting data, we must first create (acquire) a list of all schools and subsequently assign data from all relevant databases to each school. E.g., in the case of the school portal INEKO we have adopted a list of schools from the Institute of Information and Prognosis of Education (IIPE), which falls under the Ministry of Education and is an authority which manages the network of schools in the Slovak Republic. This list of schools contains the identifiers KODSKO and for us it served as a starting point - each school under the IIPE list has a profile on our portal. The adopted list of schools that we use as a basis for profiles on our portal, and for later assignment of the collected data, will be hereinafter called "database of schools on the portal."

Note: The list of schools, which in the Slovak Republic is managed by IIPE, and which serves as the basis for our "database of schools on the portal", contains for each school in SR, in addition to its name and address, the following information: information about what type of school it is, what is the language of instruction, who is the administrator, the complete contact information (phone, e-mail, website address) and how many students it has - all these data are further adopted into our "database of schools on the portal."

A problem arises when we want to assign data from other databases to this "database of schools on the portal" (this operation is called pairing of databases) and the databases with data do not contain the KODSKO codes. Then we must rely on pairing using the names and addresses of schools - here we often run into the problem that the institutions do not record the names and addresses of schools in the same way.

As an example, we can use the database of the National Institute for Certified Educational Measurements (NICEM), which provides nationwide testing of students in schools. Since NICEM does not use the KODSKO codes, in assigning the results of national testing we must tediously pair the schools from both databases according to the name and address of a school. In the better cases, the name and address of a school corresponds between the IIPE and NICEM databases completely and without duplication, the assignment of schools in this case is trivial and the data can be assigned directly without much effort. However, often there are discrepancies in the recorded names, addresses, or both of these attributes of schools. A good example is a school, which the IIPE registers as "KŠŠ - Cirkevná ZŠ" residing at "Slov.národ.povstania 4, Šahy - Ipolyság", while NICEM registers the same school under the name "Katolícka spojená škola F. Fegyvernekiho s VJM" with the address "SNP 4, Šahy".

Taking into account that Slovakia has around 4,000 schools, even if we managed to assign data for the vast majority of schools in a pair of databases with minimal effort and there remained "only" a few hundred schools that

could not be paired due to conformity of names and addresses, there would still be a time consuming task to find all the remaining several hundred assignments of schools. In addition, there are many different forms of mismatching between the recorded names and addresses of schools in different databases. In the example of the school from Šahy, we wanted to illustrate a case where neither the record of the name of the school, nor street nor municipality are not identical. In other cases, a portion of these data may correspond, but a discrepancy may occur, for example, in the manner of record of the orientation and registration number on the street where the school is located.

If the paired database does not contain usable unique identifiers of schools, and on the other hand, contains hundreds of schools that have differently registered names and addresses of schools compared to the "database of schools on the portal", it would not be efficient enough to manually search for all the links between the elements of the two paired databases. To save time, a suitable solution is to use an algorithm that compares the consistency of all pairs of elements from both paired databases and recommends the most likely pair, for which it is only necessary to confirm their adherence (no need to manually search for potential pairs). It is possible to adapt such an algorithm, or to prepare your own. In the case of the INEKO school portal we preferred the second option, because it seemed useful to have an algorithm customized to Slovak facts (according to the form of names and addresses of the schools included in paired databases, according to the usual nature of discrepancies, etc.). This algorithm will go over all the elements in a paired database, and for each element it primarily searches for the element of the "database of schools on the portal", which should have the same name and address, and the assignment must be unambiguous ("one to one"). The algorithm first checks which elements can be assigned directly. In case of the elements that could not be assigned directly, the algorithm continues to look for potential adepts from the "database of schools on the portal" from the given municipality, in a way that for each school in the municipality determines the rate of correlation according to the correlation of the school's name, street and orientation or registration number. The output of the algorithm is that for every element of the paired database it offers the best candidate for a partner from the "database of schools on the portal" and according to the setting also a number of other alternative adepts, while each adept is given a rating, which quantifies the correlation between the pair of compared elements. Subsequently, it's necessary that a person responsible for the pairing of data goes over all elements from the paired database that could not be assigned automatically, and checks whether the algorithm offered and found a suitable partner from the "database of schools on the portal" or not. If not, it's necessary to assess each situation individually and determine whether it is possible to find a partner manually or whether the relevant partner is objectively not present in the second database. It also can occur that a school is mentioned multiple times in the paired database and there are at least two assignments to one element from the "database of schools on the portal" - in this case it's necessary to assess the situation individually again, whether the pairing should use only one of the entries for the given school, or, for example, add and average out all data.

A range of unforeseen complications can always occur when pairing databases, e.g. problems with accents, typos, wrong data format, data errors and the like. The treatment of certain indicators must be therefore often modified so that errors are eliminated and the amount of successfully associated data is as large as possible.

Note: The source code of the algorithm that helps with the pairing of databases, which was prepared by INEKO according to their needs, is attached for the sake of demonstration in the annex to this manual. It is a code in Visual Basic, which links to the corresponding workbook in Microsoft Excel format. The provided source code covers a number of operations carried out in the pairing of databases and assignment of data, not only the search of adepts

for pairs between database elements, but also functions for transfer and partial processing of data, etc. The actual source code is not built universally - it is largely focused on the data allocation conditions in the Slovak Republic. The author of this manual does not expect that it will be applicable to users in other countries. The objective is rather to provide inspiration, if necessary, or enable the use of smaller pieces of the source code. The accompanying workbook in Microsoft Excel format will be provided upon request.

School portal - processing of data, basic principles of methodology and calculation of ratings

In order to neatly store all the collected data in one place and at the same time allow it to be systematically processed further, we recommend creating a central database of all the data above the "database of schools on the portal". Specifically, we recommend maintaining this central database in, for example, one large table, where each row represents one school (i.e. the number of rows in the table will be equal to the number of all schools on the portal) and columns for each school will sequentially contain the following data (if available):

- ID of the school (the order in the table; at n schools these are sequential numbers from 1 to n)
- Name of school
- Address of school (ideally divided into multiple columns for street, municipality, postal code, district, region, GPS coordinates)
- Universal school identifier (the code list can be taken from another institution)
- Any additional identifiers as needed (e.g. code lists of other institutions)
- All the characteristics of a school (e.g. type, language of instruction, specialization, etc.; each characteristic in one column)
- All contact information of a school
- All (raw / unprocessed) data for the oldest available (school) year; each indicator takes at least one column, in some cases more, for example in the case of a ratio indicator, say, if we measure the proportion of teachers using ICT regularly in teaching, it is useful to know not only the proportion itself, but also the number of teachers using ICT and the total number of teachers - this information must be maintained if we want to display these details on the portal later (not only the proportion itself)
- Then all the raw data again, but for the next (school) year up to the latest available

If we have successfully paired the necessary data sources with the "database of schools on the portal", and prepared the above-described central database of data, to which we transferred the data from the individual paired databases, we can start with the processing of data.

At first, if necessary, we complete the basic data processing - create ratio pointers (e.g. resources per student in Euros), or normalize the data (e.g. if a country is carrying out tests with the maximum obtainable score of 150, it is recommended to rescale the results to the interval of 0 to 100 percent). If possible, we prefer an adaptation of pointers where a higher achieved value represents a better result. It is not an obstacle if resulting pointers are in completely different units - that is, if some are in percentages, others in Euros per students or only in nominal terms (e.g. a number of competitions in which the school is involved).

The prepared indicators will appear on the portal, so they should be sorted as suitably as possible. The aim is to display the most significant pointers first, but the most important is their logical separation according to their affiliation (e.g. first the results of the leaving exams for all subjects, then the results of all types of inspections, and then another category of pointers, etc.). The order should be decided upon before data processing, and used during

processing so the data after processing is included in databases in the order in which it will be later displayed on the portal.

Subsequently, if we have completed individual pointers, we need to calculate percentiles and ratings in relation to them. Without them, the user may be unable to simply assess whether the selected schools achieved an above-average or below-average outcome within a selected pointer. Percentiles and ratings also have the property that they do not preserve the original unit of measurement (it does not matter whether the original pointer was e.g. in points or Euros per student). Because of this property, it will be possible to combine the adjusted pointers - namely, create an overall rating as a weighted average of several sub-ratings (for individual pointers).

Percentiles have the disadvantage that they neglect the precise differences between the results of individual schools - they simply track how many schools achieved a better result and how many schools a worse result than the selected school. The advantage is that the information regarding the percentile of the school is very easy to interpret. An user of the portal can quickly get an idea of the ranking of a school within the given aspect compared to other schools. The advantage is also that the percentiles can be easily calculated e.g. by a built-in function PERCENTRANK () in MS Excel.

The advantage of the rating is that it can retain information about how far apart the schools are in terms of results - it depends on what methodology we choose. The rating will be probably more difficult to interpret than a simple percentile. The basic approach is via the rescaling of values within the given pointer so the school with the worst result (among the schools of this type in the given school year) is assigned a rating of zero (0 points) and the school with the best result the maximum rating, thus e.g. 10 points (for simplicity, we will be using ratings in the range from 0 to 10 points). All other schools of the given type will be assigned a rating for the given pointer using the formula: $\text{rating} = (\text{pointer value for a given school} - \text{the worst value reached among the schools of this type in a given year}) / (\text{best} - \text{worst value achieved among the schools of this type in a given year})$. It's a linear transformation of the original values to a new interval, in this case the interval from 0 to 10 points. The disadvantage of this approach is the high sensitivity to fluctuations of limit values - e.g. if in some year a school reaches a very bad result and if this bad result is isolated, then all the other schools obtain a relatively high rating (the average rating will be substantially higher than 5 points). It could therefore easily occur that a school, which is achieving average (let's say median) results every year, will have a significantly above average rating in one year, and a significantly below average rating in the following year. As this property is not desired, it is recommended to improve the methodology to mostly eliminate this property, for example by defining a "notional minimum and maximum" (the reached values that deserve to be allocated a minimum and maximum rating) by other means, say through the 5th and 95th percentile. In the above formula we therefore replace the worst value among schools of this type in that year by the reached value of such a school of this type in the given year that only 5% of all schools of this type in that year achieved a worse result. Similarly, the best value among schools of this type in that year will be replaced by the achieved value of such a school of that type in that year, so that only 5% of all schools of that type achieved a better result (or 95% of schools achieved a worse result). Another possibility of developing the methodology is that for the determination of "notional minimum and maximum" we will only consider schools that have some minimum sample for which the given pointer was created (i.e. for example, we will search for 5th and 95th percentiles only among schools that had at least 20 students tested). All schools which should get a rating of less than 0 points according to this approach will be assigned a rating of zero (0 points). Similarly, schools that should get a higher rating than 10 points will be assigned a rating of 10 points.

We emphasize that in both cases it is important to count (percentiles and ratings) only on the basis of data for the same type of school. Otherwise, it would not ensure the comparability of schools based on the calculated percentiles or ratings.

Ratings for individual pointers prepared in this way can be further aggregated and thus create ratings for the entire groups of pointers and, ultimately, the overall rating for the whole school. We recommend calculating these ratings as weighted averages of sub-ratings, while the weight should reflect the significance (informative value) of the included pointers. For example, in the case of the Slovak portal, only about half of the pointers are included in the calculation of the overall rating of a school. The pointers that are more informative in character, and do not have the sufficient capacity to testify about the quality of schools, should not be included in the formula for calculating the overall rating of a school. When calculating the overall rating, we also recommend omitting those pointers, where values are available only for a small number of schools (e.g. for less than $\frac{3}{4}$ s of all schools). Another enhancement of the methodology is that, within the Slovak portal, we calculate ratings based on data for the past 4 years, and the most recent figures have a higher weight than older ones - the goal is to minimize any single random fluctuations in the values achieved by schools, and to increase the data sample from which the ratings are calculated. In general, the ratings are not calculated for schools where there are only small samples available (we set the minimum sample to 80 students for the past 4 years).

School portal - methodology and calculation of ratings used within the portal in SR

In this part of the manual we provide the specific procedure that we used in preparation of the Slovak portal. The methodology was selected to meet the conditions specified in the previous section and to be as customized to the set of data available in SR as possible.

For the purposes of the portal, we were able to gather 32 pointers in total, for which data were available and which were in line with the conditions set out in the previous sections of the manual. The pointers are divided into 9 groups - Leaving examinations, Unemployment of graduates, University admissions, Inspection, Extraordinary results, Participation in competitions, Teaching staff and Financial resources. The pointers are as follows:

- Leaving examinations
 - Slovak language
 - Mathematics
 - Hungarian language
 - Slovak language and Slovak literature
 - English language B1
 - English language B2
 - German language B1
 - German language B2
- Testing 9
 - Slovak language
 - Mathematics
 - Hungarian language
 - Slovak language and Slovak literature
- Unemployment of graduates
 - Unemployment of graduates
 - Regional application
- University admission
 - Proportion of graduates admitted to universities
 - University admission success
- Inspection
 - School management
 - Conditions of training and education
 - Educational process
 - Slovak language - 9th grade

- Natural science - 4th grade
- Physics - 9th grade
- Biology - 9th grade
- Extraordinary results
 - Extraordinary results of students
- Participations in competitions
 - Number of participations in competitions
 - Achieved results in the competition "Komparo" - Mathematics
 - Achieved results in the competition "Komparo" - Slovak language
- Teaching staff
 - Number of teachers per 100 students
 - Utilization of ICT
 - Proportion of qualified teachers
- Financial resources
 - Resources per student
 - School's own resources

Note: Not all pointers are available for both primary and secondary schools - for example, Leaving examinations are conducted only in secondary schools and Testing 9 only in primary schools. In addition, not all presented pointers are included in the overall rating of a school.

At first, we will calculate the ratings for individual pointers for individual years. The rating of a selected pointer for a given school and year is determined as $\max(0; \min(10; 10 * (\text{pointer value for a given school and year} - 5\text{th percentile from values achieved among schools of this type in the given year}) / (95\text{th percentile from values achieved among schools of this type in the given year} - 5\text{th percentile from values achieved among schools of this type in the given year})))$. Percentiles are determined only on the basis of those schools that had a sufficient sample of students (at least 20).

We subsequently aggregate the ratings for individual years for a given pointer to get the rating for the given pointer for the entire period (overall rating for the last 4 years). The basic principle in calculating the ratings for multiple years is always the same – a rating is created as a weighted average of sub-ratings for the most recent four years with weights of 0.4 for the latest year, 0.3 for the previous year, 0.2 for the year before and 0.1 for the oldest year.

According to the above principle, we create weighted averages of sub-ratings for individual years and thus get ratings for individual pointers. Then, using the appropriate weights we aggregate them further to the level of ratings for groups of pointers. It applies that the rating of the groups of pointers is given as a weighted average of ratings of all pointers in the group of pointers.

Calculation of ratings of groups of pointers from the ratings of individual pointers (with corresponding weights):

- Testing 9
 - 50 % Slovak language

- 50 % Mathematics

The criterion for assigning rating: at least 80 students tested in Mathematics and Slovak language for the last 4 years

Note: For schools with Hungarian language of instruction are, instead of the pointer "Slovak language", taken into account two pointers "Slovak language and literature" and "Hungarian language" with half the weight

- Leaving examinations

- 16.6 % Slovak language
- 16.6 % Mathematics
- 16.6 % English language B1
- 16.6 % English language B2
- 16.6 % German language B1
- 16.6 % German language B2

The criterion for assigning rating: at least 80 students with leaving examination in Slovak language and 80 in foreign languages for the last 4 years

Note: For schools with Hungarian language of instruction are, instead of the pointer "Slovak language", taken into account two pointers "Slovak language and literature" and "Hungarian language" with half the weight

- Unemployment of graduates

- 0 % Unemployment of graduates
- 100 % Regional application of graduates

The criterion for assigning rating: at least 80 graduates in total for the last 4 years

- University admission

- 50 % Proportion of graduates admitted to universities
- 50 % University admission success

The criterion for assigning rating: at least 40 graduates or 20 graduates admitted to university for the last 4 years

- Inspection

- 14.3 % School management
- 14.3 % Conditions of training and education
- 14.3 % Educational process
- 14.3 % Slovak language - 9th grade
- 14.3 % Natural science - 4th grade
- 14.3 % Physics - 9th grade
- 14.3 % Biology - 9th grade

The criterion for assigning rating: none

- Extraordinary results
 - 100 % Extraordinary results of students
The criterion for assigning rating: at least 80 students in total for the last 4 years
- Participations in competitions
 - 33,3 % Number of participations in competitions
 - 33,3 % Achieved results in the competition "Komparo" - Mathematics
 - 33,3 % Achieved results in the competition "Komparo" - Slovak language
The criterion for assigning rating: none
- Teaching staff
 - 60 % Number of teachers per 100 students
 - 20 % Utilization of ICT
 - 20 % Proportion of qualified teachers
The criterion for assigning rating: at least 40 teachers in total for the last 4 years
- Financial resources
 - 80 % Resources per student
 - 20 % School's own resources
The criterion for assigning rating: at least 80 students in total for the last 4 years

The procedure of calculating the ratings for individual groups of pointers is therefore very similar for all of the groups of pointers and all types of schools. The overall rating of a school ultimately always arises as a weighted average of ratings for the selected groups of pointers.

In case of primary schools and special primary schools, the calculation of the overall school rating includes pointers from the groups "Testing 9" and "Extraordinary results" in the case of secondary vocational schools, special secondary schools and gymnasiums, they include pointers from the groups "Leaving examinations", "Extraordinary results" and "Unemployment of graduates." The portal calculates the ratings for other groups, but those ratings do not enter into the overall school ratings, since they may not provide sufficient information about the results of the students.

Calculation of the overall assessment (rating) of a school from the ratings of groups of pointers (with corresponding weights):

- Primary schools
 - 80 % Testing 9
 - 20 % Extraordinary results
- Secondary vocational schools
 - 40 % Leaving examinations
 - 10 % Extraordinary results
 - 50 % Unemployment of graduates
- Gymnasiums

- 60 % Leaving examinations
- 15 % Extraordinary results
- 25 % Unemployment of graduates

The overall rating will be assigned only to those schools that have a sufficient number of students or graduates:

- Primary schools and special primary schools: at least 80 students tested in the Slovak language (or the Slovak language and literature and the Hungarian language in schools with instruction in the Hungarian language) and mathematics for the past 4 years
- Secondary vocational schools, special secondary schools and gymnasiums: at least 80 students with leaving examination in the Slovak language (or the Slovak language and literature and the Hungarian language in schools with instruction in the Hungarian language), 80 students with leaving examination in foreign language and 80 graduates
- Primary art schools and conservatories do not receive ratings

Please note: The calculation of ratings for certain pointers and groups of pointers is slightly different from the basic principle. For the pointers that enter into the calculation of the overall rating of schools, the weight of individual years (0.1, 0.2, 0.3, 0.4) is multiplied by the number of students, on the basis of which the value was obtained. For example, if a small school in the school years 2009/10, 2010/11, 2011/12 and 2012/13 had successively 40, 35, 30 and 25 graduates each year, the weights for calculating the assessment for the Unemployment of graduates will be successively 4/30, 7/30, 09/30 and 10/30 (modified weights arose as a product of the original weights and the number of graduates divided by a constant at which the sum of adjusted weights is equal to 1). The same applies for the results of Leaving examinations (number of participants), Testing 9 (number of students tested) and for Extraordinary results (number of students). The aim is to take into account the different size of samples.

Please note: For the pointers that are not limited from above or below (e.g. number of teachers per 100 students or resources per student), prior to the calculation of rating the extreme (unexpected) values must be eliminated - the values that are either greater than the third quartile + 3 * inter-quartile range, or smaller than the first quartile - 3 * inter-quartile range

Note: The rating is always a number from 0 to 10 and the higher the number achieved the better the assessment. If a rating of a pointer is not available for any relevant type of school in a given year (e.g. Unemployment of graduates in 2013/14, which will be based on the methodology available in October 2015) the weight of the pointer in a given year shall be 0.

Calculated overall ratings of schools are complemented by verbal assessment in the profiles of schools on the portal (verbal assessment applies only to the overall school rating and does not apply to the rating of individual areas):

- 0.00 - 1.99 school with very bad results
- 2.00 - 2.99 school with bad results
- 3.00 - 3.99 school with below average results
- 4.00 - 4.99 school with average results
- 5.00 - 5.99 school with good results
- 6.00 - 6.99 school with very good results

- 7.00 - 7.99 school with excellent results
- 8.00 - 10.00 school with exceptional results

After the completion of the entire procedure, the following data is ready to be uploaded to the portal:

- "Database of schools on the portal", including the characteristics of schools and contact information
- The list of pointers along with their properties (names of pointers, units, assignment to groups of pointers)
- Values of pointers for individual schools and individual years (possibly along with supporting data, such as sample size, etc.).
- Percentiles corresponding to the pointers for individual schools and for individual years
- Ratings pertaining to the groups of pointers for individual schools and for individual periods (e.g. four-year periods)
- Overall ratings for individual schools and for individual periods (e.g. four-year periods)

Upload the above data may be complemented with comments received from representatives of individual schools. If a school representative wishes to comment on the published data or highlight some of the achievements that cannot be captured by the published measurable pointers, he/she can prepare a comment in the form of text, which can also be uploaded to the portal and published under the profile of the school.

Local government management portal - selection and acquisition of data

In the case of the local government management portal the aim was to collect all relevant accounting information on the management of local governments, in a way to enable the monitoring of fulfilment of existing legal criteria for management, and also assess the financial health of individual local governments (according to own methodology).

Note: Local governments in the SR include a total of 2,930 municipalities (including cities and urban areas) and 8 higher territorial units (HTU).

In the case of the local government management portal there is an advantage in the fact that in the Slovak Republic all relevant information is administered by the Data Centre, which reports directly to the Ministry of Finance. We requested the information from the Ministry of Finance under the Act on Free Access to Information. The data is annually submitted to the Data Centre by municipalities via statistical reports.

For the needs of the portal we receive the following financial pointers for each local government from the Ministry of Finance, which we further process to financial stability pointers and the overall assessment of financial health:

- Income of the current municipal (HTU) budget per calendar year; it's the really achieved (not budgeted) income and the data includes budgetary organizations within the scope of the municipality (HTU)
- Capital income of the municipal (HTU) budget per calendar year; it's the really achieved (not budgeted) income and the data includes budgetary organizations within the scope of the municipality (HTU)
- Expenses of the current municipal (HTU) budget per calendar year; it's the really achieved (not budgeted) expenses and the data includes budgetary organizations within the scope of the municipality (HTU)
- Capital expenses of the current municipal (HTU) budget per calendar year; it's the really achieved (not budgeted) expenses and the data includes budgetary organizations within the scope of the municipality (HTU)
- Total volume of loans of the municipality (HTU) from the Housing Development Fund (including interest payable); the data includes budgetary organizations within the scope of the municipality (HTU)
- Total overdue liabilities of the municipality (HTU); the data includes budgetary organizations within the scope of the municipality (HTU); this is the status of the liabilities as of December 31 of the given year
- Total overdue liabilities of the municipality (HTU) unpaid for 60 and more days after the due date; the data includes budgetary organizations within the scope of the municipality (HTU); this is the status of the liabilities as of December 31 of the given year; note: if the municipality (HTU) has not paid a recognized liability within 60 days from the day of maturity, it fulfils one of the two conditions for the obligation to draw up a proposal for the introduction of a recovery regime
- Total volume of fixed assets of the municipality (HTU); the data includes budgetary organizations within the scope of the municipality (HTU); this is the status of the assets as of December 31 of the given year

- Total volume of short-term receivables of the municipality (HTU); the data includes budgetary organizations within the scope of the municipality (HTU); this is the status of the short-term receivables as of December 31 of the given year
- Total volume of funds on the financial accounts of the municipality (HTU); the data includes budgetary organizations within the scope of the municipality (HTU); this is the status of the accounts as of December 31 of the given year
- Total volume of long-term liabilities of the municipality (HTU); the data includes budgetary organizations within the scope of the municipality (HTU); this is the status of the long-term liabilities as of December 31 of the given year
- Total volume of short-term liabilities of the municipality (HTU); the data includes budgetary organizations within the scope of the municipality (HTU); this is the status of the short-term liabilities as of December 31 of the given year
- Total volume of bank loans of the municipality (HTU); the data includes budgetary organizations within the scope of the municipality (HTU); this is the volume of the loans as of December 31 of the given year
- The number of residents of the municipality (HTU) as of January 1 of the given year
- Total expenses of the municipality (HTU) for payment of principal on loans; it's the really achieved (not budgeted) expenses and the data includes budgetary organizations within the scope of the municipality (HTU)
- Total amount of interest payments of the municipality (HTU); it's the really paid (not budgeted) expenses and the data includes budgetary organizations within the scope of the municipality (HTU)
- The achieved economic result of the municipality (HTU) for the accounting period (calendar year); the data includes budgetary organizations within the scope of the municipality (HTU)
- The total amount of the debt according to legal definition as of December 31 of the given year
- Additional data (statute, e-mail address)

Note: Thanks to the fact that the collection of data from all municipalities is provided by the same institution, the potential for any discrepancies in the data reporting methodology of individual municipalities had decreased. Still, in the course of the project we encountered several cases where a municipality had clearly incorrectly recorded data in the database of the Ministry of Finance. It was ultimately discovered in all cases that there had been incorrect reporting of information from municipalities. It seems that now that the public can easily access this data, the error rate has significantly decreased (probably due to the fact that municipalities know that they must be more careful when reporting data, because they are subject to public scrutiny to a much larger extent).

Note: In the case of local governments, we get all the underlying data from one institution, so there is practically no problem with the pairing of databases. It is true that the Ministry of Finance records the names of municipalities in two ways (depending on two different groups of databases), but the pairing of these two sets of names represents a much smaller problem than in the case of schools - therefore we will not discuss it in particular.

Note: As in the case of schools, in the case of municipalities we publish a list of characteristics (statute, number of residents) and contact information (e-mail address, GPS coordinates) in relation to the individual municipalities.

Local government management portal - processing data, design of methodology and calculation of the financial health of local governments

Based on the acquired pointers, we prepare a total of 15 indicators of financial stability (in most cases they are ratio indicators). The aim of some of them is to cover all Slovak legally defined criteria for management; the rest are indicators that also have good informative value when assessing the management of local governments. Together with indicators we also provide their description, and if we assign a rating also its assignment process:

- Basic balance of the municipality (HTU) is calculated as $(\text{current income} + \text{capital income} - \text{current expenditure} - \text{capital expenditures}) / (\text{current income} + \text{capital income})$. The pointer indicates whether the local government managed with surplus or deficit in terms of its current and capital account as a whole. It informs about the fact how the local government is able to cover its current expenses (operation of the local government) and capital expenses (asset enhancement) from its current income (e.g. tax income) and capital income (e.g. sale of assets). If the basic balance of the municipality (HTU) is at a level below or equal to -15%, the assigned rating is 0. The assigned rating is linearly rising with the improving basic balance due to the decreasing risk for the municipality (HTU). For a basic balance at the level of 0% the rating equals 3 (middle of the scale) and for a value equal to or greater than +15% the assigned rating is 6. In other words, if the basic balance is positive, the municipality (HTU) acquires a rating greater than 3 (according to the results achieved). Otherwise, it is assigned a rating in the lower half of the scale.
- Debt service of the municipality (HTU) is calculated as $(\text{payment of principal} + \text{interest payments}) / (\text{current income for the previous year})$. This pointer informs about the expenses of the local government in connection with the debt service. Under the law, the expenses for payment of principal and interest should not include the on-time early repayment. Note: The legal limit is 25%; if the municipality exceeds this limit, it cannot accept other repayable sources of funding under the law. If the debt service of the municipality (HTU) is at the level of 0%, the assigned rating is 6. The assigned rating is linearly rising with the increasing debt service due to the increasing risk for the municipality (HTU). For a debt service at the level of 25% the rating equals 3 (middle of the scale) and for a value equal to or greater than 50% the assigned rating is 0.
- Data on the total debt of municipalities (HTU) are, from 2012 onwards, collected from the MoF. It's a so called "statutory criterion", i.e. the debt according to Section 17 par. 8) of Act 583/2004 Coll. (not counted are liabilities from loans provided by former state funds and the State Housing Development Fund, and also the obligations of reimbursable financing resources taken to ensure the pre-financing of EU funds). The debt is reported in relation to current income for the previous year and according to the law it should not exceed 60%. Data on compliance with statutory debt criteria are not available for municipalities and higher territorial units for 2006-2011 (the MoF does not have the data on the reimbursable sources of financing taken to ensure the pre-financing of EU funds). The total debt of municipalities (HTU) is thus, prior to 2011, approximated based on available data (bank loans and borrowings + long-term liabilities - loans from the

State Housing Development Fund) / (current income for the previous year). Please note: The reported debt may not include the entire indebtedness of municipalities (HTU) in the case that the debt is "hiding" e.g. through subsidiary companies, PPP projects and supplier credits. If the total debt of the municipality (HTU) is at the level of 0%, the assigned rating is 6. The assigned rating is linearly rising with the increasing debt due to the increasing risk for the municipality (HTU). For a debt at the level of 60% the rating equals 3 (middle of the scale) and for a value equal to or greater than 120% the assigned rating is 0.

- The liabilities of the municipalities (HTU) that are at least 60 days overdue are calculated as (liabilities overdue for 60 days or more) / (current income for the previous year). If the municipality (HTU) has not paid a recognized liability within 60 days from the day of maturity, it fulfils one of the two conditions for the obligation to draw up a proposal for the introduction of a recovery regime. If the amount of debt of municipality (HTU) overdue for at least 60 days is exactly 0 Euros or 0 % the assigned rating is 6. If the amount of these liabilities in proportion to income represents a value between 0% and 3%, the assigned rating is between 3 and 0 (linear) and if the value is at a level equal to or greater than 3% the rating is 0.
- Immediate liquidity of the municipality (HTU) is calculated as (financial accounts) / (short-term liabilities). The pointer informs about the extent to which there are available funds in financial accounts of local governments sufficient for payment of short-term liabilities. If the immediate liquidity of the municipality (HTU) is at the level of 0%, the assigned rating is 0. The assigned rating is linearly rising with the increasing immediate liquidity due to the decreasing risk for the municipality (HTU). For immediate liquidity at the level of 100% the rating equals 3 (middle of the scale) and for a value equal to or greater than 200% the assigned rating is 6. In other words, if the amount of funds in the financial accounts of the municipality (HTU) is greater than the amount of its short-term liabilities, it gains a rating higher than 3 (according to the size of this ratio). Otherwise, it is assigned a rating from the lower half of the scale.
- Current account balance of the municipality (HTU) is calculated as (current income - current expenses) / (current income). The pointer indicates whether the local government managed with surplus or deficit in terms of current account. It informs about how the local government is able to cover its current expenses (operation of the local government) from its current income (e.g. tax income).
- Capital account balance of the municipality (HTU) is calculated as (capital income - capital expenses) / (capital income). The pointer indicates whether the local government managed with surplus or deficit in terms of capital account. It informs about how the local government is able to cover its capital expenses (enhancement of assets) from its capital income (e.g. sale of assets).
- The amount of overdue liabilities (HTU) to income is calculated as (overdue liabilities) / (current income for the previous year). The total amount of overdue liabilities of the municipality (HTU) exceeds 15% of the actual current income of the municipality for the previous financial year; it fulfils one of the two conditions for the obligation to draw up a proposal for the introduction of a recovery regime.
- Net assets of the municipality (HTU) is, for the years 2006-2011, calculated as (fixed assets + financial accounts - (bank loans and borrowings + long-term liabilities - loans from the State Housing Development Fund)) / (current income for the previous year). From 2012 are also available the data on the amount of liabilities from returnable sources of financing taken to ensure the pre-financing of EU funds, which are, under the law, not included into the total debt of municipalities (HTU). Therefore, from 2012 onwards, the net assets of the municipality (HTU) are calculated as (fixed assets + financial accounts - (bank loans and borrowings + long-term liabilities - loans from the State Housing Development Fund - pre-financing

liabilities from Euro funds)) / (current income for the previous year). The pointer informs about the amount of local government assets (net of the debt) relative to its current income per year.

- Quick liquidity of the municipality (HTU) is calculated as (financial accounts + short term receivables) / (short-term liabilities). The pointer informs about the extent to which there are available liquidity resources of local governments sufficient for payment of short-term liabilities. As the liquidity resources can be considered in addition to funds in financial accounts also the total amount of short-term receivables of the local government.
- The volume of loans of the municipality (HTU) from the Housing Development Fund in proportion to income is calculated as (loans from State Housing Development Fund) / (current income of the previous year), and the loans from the State Housing Development Fund also include the liabilities resulting from interest.
- Basic balance of the municipality (HTU) per 1 resident is calculated as (current income + capital income - current expenditure - capital expenditures) / (number of residents at the beginning of the year). Basic balance indicates whether the local government managed with surplus or deficit in terms of current and capital account as a whole. It informs about how the local government is able to cover its current expenses (operation of the local government) and capital expenses (asset enhancement) from its current income (e.g. tax income) and capital income (e.g. sale of assets).
- Debt of the municipality (HTU) per one resident for 2006-2011 is calculated as (bank loans and borrowings + long-term liabilities - loans from the State Housing Development Fund) / (number of residents at the beginning of the year). Starting from 2012, it is calculated on the basis of data from the MoF on the total debt of municipalities (HTU) in proportion to current income for the previous year (data are then calculated per 1 resident).
- Net assets of the municipality (HTU) per one resident for 2006-2011 is calculated as (fixed assets + financial accounts - (bank loans + long-term liabilities - loans from the State Housing Development Fund) / (number of residents at the beginning of the year). From 2012 are also available the data on the amount of liabilities from returnable sources of financing taken to ensure the pre-financing of EU funds, which are, under the law, not included into the total debt of municipalities (HTU). Therefore, from 2012 onwards, the net assets of the municipality (HTU) per 1 resident are calculated as (fixed assets + financial accounts - (bank loans and borrowings + long-term liabilities - loans from the State Housing Development Fund - pre-financing liabilities from Euro funds)) / (number of residents at the beginning of the year). The pointer informs about the amount of local government assets (net of the debt) calculated per 1 resident.
- Economic result of the municipality (HTU) per 1 resident is calculated as (economic result for the current period) / (number of residents at the beginning of the year).

Another extension is the assessment of financial health for five key areas (total debt, debt service, liabilities at least 60 days overdue, immediate liquidity and basic balance). The financial health is one number between 0 (worst) and 6 (best); it is therefore the same interval as in the rating for indicators of financial stability. The procedure of calculating the rating for each area is as follows:

- The score for "Total debt" within the financial health is equal to the score of the indicator of financial stability Total debt for the latest year available.

- The score for the "Debt service" within financial health is calculated in two steps. The first step is calculating a weighted average of the debt service values for the latest available year and the three preceding years, with the successive weights 4, 3, 2, 1. If there is no score available for some of the years, that year is not taken into account. The second step is to calculate a score for this average value of debt service in the interval of 0 to 6, exactly according to the methodology for calculating the score for the financial stability indicator Debt service.
- The score for "Liabilities overdue for at least 60 days" within the financial health is calculated as a weighted average of the financial stability indicator Liabilities overdue for at least 60 days for the latest available year and the three preceding years, with the successive weights 4, 3, 2, 1. If there is no score available for some of the years, that year is not taken into account.
- The score for "Immediate liquidity" within the financial health is calculated as a weighted average of the financial stability indicator immediate liquidity for the latest available year and the three preceding years, with the successive weights 4, 3, 2, 1. If there is no score available for some of the years, that year is not taken into account.
- The score for the "Basic balance" within financial health is calculated in two steps. The first step is calculating a weighted average of the basic balance values for the latest available year and the three preceding years, with the successive weights 4, 3, 2, 1. If there is no score available for some of the years, that year is not taken into account. The second step is to calculate a score for this average value of basic balance in the interval of 0 to 6, exactly according to the methodology for calculating the score for the financial stability indicator Basic balance.

Finally, the overall rating of the financial health of municipalities / HTU is calculated as a weighted average of the scores achieved by each of the five aforementioned components. All components except the first have a weight of 1, the first component - total debt - has a weight of 2. The financial health informs about the extent to which management is sustainable and whether the management of a particular town, municipality or HTU causes problems.

The overall rating is connected to the following verbal assessment of local governments:

- 0.00 - 2.99 insufficient financial health
- 3.00 - 3.99 sufficient financial health
- 4.00 - 4.99 good financial health
- 5.00 - 6.00 excellent financial health

The data prepared in this way for each local government is ready to be uploaded on the portal. Thanks to this data, users have in one place access to comprehensive information about local government - administrative and geographical data, input financial pointers, indicators of financial stability as well as financial health.