|  |
| --- |
| 1. **Name of the challenge** *(short, powerful and inspiring description):*

 **Intelligent importer of public datasets** |
| 1. **Context*:*** *(what is the background information behind the challenge, what is the state of the art of the sectors, the role of the organization in this context, the target group to whom the solution need to be addressed, etc.)*

 To improve the existing tool for automated unification of dataset schemes in the public administration. The focus of the challenge is enhancement of tool robustness, in order it could process as wide as possible spectrum of dataset cases. The solvers should develop software that can be used in the whole public sector. They will gain experience in transformation and linking of data using the AI methods. You may get a chance to cooperate on national standards for all IT projects. |
| 1. **Problem:** (*What i*s *the problem that needs to be solved, why is important to solve, impact of this problem in the close future, impact of the problem on local or international area)*

It is in the interest of the state authorities to publish its data in the best machine readable form, in particular in a unified scheme in order to enable automated interlinking of individual datasets. The central point of publication of these data is the information portal data.gov.sk. At the same time is in use also a unified data scheme (Central model of public administration data), that describes linkages by semantic relationships. However, there persists a problem of unification of individual datasets. Currently they exist in various formats and even with unknown structure of content (semantics) that makes impossible their linear linkages, e.g. the attribute of “firm” may occur in another dataset as “company”, etc. The consequence is huge amount of time and money is repeatedly spent on data integration. The administrator of these datasets, usually officials, have no technical capabilities to harmonize their datasets with the central scheme, thus making their output data difficult to be used by a different public administration subject and commercial sector as well. The tool shall transform the dataset to meet the requirements of central scheme for, interlinking of individual published datasets and provide to the users a very user-friendly interface.1. **Additional info (for internal use):**  *(what is expected to be delivered by the team (idea/concept/prototype), what are the specific tools & instruments that shall be used (e. g. Programing language etc.), what are the asset (as knowledge, materials) will be given to the team*

The project will require from the solvers work with text similarities regular expressions, data retrieval, graphs, and machine learning in text processing. Solvers will be faced with some UX tasks as well, since on the one hand the user interface of the tool must be within the capabilities of an ordinary clerk without any special IT skills and on the other hand it must guide the user so that the output from the tool achieves excellent results. |
| 1. **Skills of the team (for internal use):** *what specific skills shall the team have in order to address the challenge*

Obligatory skills: Java (backend), Git, CI, Docker. Recommended skills: AngularJS, ElasticSearch. Useful knowledge fields: Information retrieval, Knowledge discovery, Neuron networks, Statistical methods of experiment evaluation. |
| **5. About the Seeker:****The Office of Vice Prime Minister of Slovak Republic, Section on Investments and Informatisation.** The mentor on behalf of the Seeker : Ing. Marek Šurek acting in several standardisation commissions in this Office. He will be helpful in creation and description of processes, that can be trasformed into a standard obligatory for information systems in the public administration. |