Animal welfare has received much attention and good welfare was shown to increase farm profitability. However, measuring animal welfare on farm is challenging as it requires animal-based measures typically gathered during an on-farm visit. Herd performance data are recorded routinely worldwide and may provide insight into specific aspects of animal welfare. The objective of this study was to develop a tool that evaluates remotely the herd welfare status based on a multi-criteria evaluation approach using pre-recorded herd data. The tool created includes 14 indicators covering important aspects of animal welfare (longevity, nutrition, production, management, health, and reproduction) over the life span of a dairy cow. Data were normalized to percentile ranks and aggregated to a composite welfare index (CWI), scaled to percentile ranks, with equal weights among indicators. The percentile rank of each indicator aimed at evaluating specific herd strengths and weaknesses, and that of the CWI at ranking the herd within a population of herds, with a low CWI indicating potential welfare issues. The CWI was generated for 4,003 Canadian herds in 2017 and 2018. Potential sources of variation in the generation of the CWI and the robustness of the CWI over time between 2017 and 2018 were evaluated. The results showed that this approach allows screening and comparing the herd welfare status among herds through their CWI. The ranking of herds fluctuated to some extent, which is an inherent issue of any indexing method. However, the CWI fluctuated less for low-ranking herds with potentially greater welfare issues (<10th percentile; CWI of 0.07 ± 0.061; mean ± standard deviation) and for high-ranking herds with potentially few welfare issues (>90th percentile; CWI of 0.94 ± 0.062) as compared to average herds ranking between the 25th and 75th percentile (CWI of 0.49 ± 0.157). These results suggest that the proposed approach might be useful to monitor the top and bottom welfare herds. In-depth analysis revealed that the stability of CWI was mainly influenced by the presence of extreme values for the individual welfare indicators. Careful removal of extreme values might improve the overall robustness of the CWI to some extent. A comparison of the CWI between 2017 and 2018 revealed a mean absolute difference (MAD) of 0.135 ± 0.1163. These results suggest that 51% of the herds fluctuated by more than 10 percentile ranks, and 72% of the herds by more than 5 percentile ranks. The MAD of individual welfare indicators ranged from 0.003 to 0.271, and revealed potential differences in sensitivity among indicators. With this tool, producers can remotely benchmark their own performance relative to others, adopt tailored management strategies, and monitor whether improvements put into place are effective. Advisory services can be allocated to herds at risk of poor welfare and can provide producers with opportunities for improvements by addressing specific welfare aspects.