



Reprex

Big data for small organizations.

Yes!Delft



## Big Data For Small Organizations

We bring the benefits of big data to small organizations: SMEs, NGOs, university research labs, management and policy consultants. We help our clients to get access to large enough datasets, and to build trustworthy scientific, policy, business, or AI applications with the data. Reprex grew out of CEEMID, a project that connected data and pooled surveys across 12 countries and more than 50 music industry stakeholders, creating thousands of relevant policies, statistical and scientific indicators for the music and creative industries.

Our flagship demo projects are the [Listen Local](#) ethical music recommendation system based on our [Demo Music Observatory](#) data integration and knowledge sharing platform.

👉 See our [services](#): [data curation](#), [open data access](#), [survey harmonization](#), [reproducible research](#) and validated [trustworthy AI applications](#).

📄 Download our [introduction](#).

📰 Follow [news about us](#) or the more comprehensive [Data & Lyrics](#) blog.

✉️ Contact [us](#).

**Reprex** is a reproducible research company that brings the benefits of big data to small organizations with data pooling, sharing, exploiting open sources and automatic labor-intensive tasks.

Big data creates inequalities, because only the largest corporations, government bureaucracies and scientific organizations have the human and financial resources to create large, consistent datasets. Small organizations will remain at a competitive disadvantage if they do not exploit the open data sources created by governments, replicable scientific datasets, and reproducible research innovations to keep data costs low. Because the use of AI requires particularly big datasets, they will be unable to exploit the benefits of algorithms, or worse still, algorithms will work against them.

We help our clients to get access to large enough datasets, and to build trustworthy scientific, policy, business, or AI applications with the data. Reprex grew out of **CEEMID**, a project that connected data and pooled surveys across 12 countries and more than 50 music industry stakeholders, creating thousands of relevant policies, statistical and scientific indicators for the music and creative industries. (See the Central European Music Industry Report 2020 at [ceereport2020.ceemid.eu](https://ceereport2020.ceemid.eu))

We offer services related to access to data and using data in a trustworthy way. We encourage our clients to become partners in our permanent data collection and sharing platforms, the data observatories.

- [Data Curation](#): find the best value for money data solution to your research or evaluation project.

- [Open Data](#): access to thousands of statistical, business and policy indicators in the cultural and creative industries and sustainability.
- [Survey harmonization](#): re-use already asked survey questions, use harmonized questions banks, create longitudinal and cross-sectional datasets.
- [Reproducible Research](#): automate data collections, processing, validation, corrections, documentation to save error-prone human work.
- [Trustworthy AI](#): validated, trustworthy AI applications or detect the biases of algorithms working against you.

Our flagship demo projects are the [Listen Local](#) ethical music recommendation system based on our [Demo Music Observatory](#) data integration and knowledge sharing platform. You can find our globally used [open-source statistical software](#) and [some publications](#) made with our data services on our website. We try to evangelize trustworthy AI on our [Data & Lyrics](#) blog.

## Cooperation

We are often contacted to provide access to cheap data. While we believe that access to affordable data is an important element of our services, we often find that our partners are not informed about data availability. We prefer to help the curation of the necessary data for a business, policy, or scientific project because often an imagined, expensive, or not-available data source can be easily replaced with something that is available or affordable. Consulting about data possibilities and finding the right data sources for your project is often more valuable than access to cheap data.

We prefer to make contracts that are long-term or have a long-term outlook. The secret of cheap or affordable data is continuous data collection. Data that is requested ad hoc is usually expensive or low-quality. We can give significantly better prices and eventually better services if you our clients are committed to partner up in one of our data observatories, or permanent data collection programs.

We would love to help with AI related problems, like training algorithms or validating them. With econometric and social modelling, or various business analytics solutions. We believe that small organizations are in a competitive disadvantage in this field, because they do not have the human and financial resources to invests into large enough data pools and data processing and normalization to train algorithms or good models. Our data observatories offer a data-as-service where we make the necessary investment, and your organization has continuous access to large, well-processed datasets for continuous use.

- Not convinced? Challenge us and ask us to make available a data that you cannot find in one of our data observatories.
- You can contract us for a single dataset, a small research program, if you want to, but we will encourage you to have an open mind about an ongoing relationship.

- If you cannot invest into a permanent data science capacity in your organizations, take our offer for data-as-service by joining our data observatories or reproducible-research-as-service.

### How to join our data observatories?

We are flexible in the terms of partnering up with our data observatories. What we ask from partners is a letter of intent in the participation of the joint activities below, and an immediate budgetary or data asset contribution when joining.

**Participate in data curation:** help observatory partners to understand data needs and potential data sources. We add value by automated processing and controls, but we need domain specific academic, business, policy expertise to assess the value of the data.

**Shared exploitation:** Use the observatory data tables, data assets, and publicly reference them. Some of the observatory assets are open for the public, others by the nature of the scientific or regulatory use, will become public after peer-review or external audit. We prioritize the embargo, business, or other interests of our observatory partners.

**Funding:** a commitment to participation in funding. We are flexible to adjust to project-based and permanent funding, and various budgetary options available for academic institutions, trade associations and collective management societies, business, and policy consultancies. We give priority for non-project-based funding because continuous data collection allows significant costs savings but requires continuous base funding.

**Validation:** In use, critically test, assess the joint data assets, and suggests improvements. Our data assets are open to many levels of validations: we give full access to the data and the code that creates the indicators, and we validated both the data and the code. We believe in peer-review and external audit, and we encourage our users to use our data in peer-reviewed or externally audited applications. We are committed to use the feedback from such external reviews promptly for continuous quality improvement.

**Seek recognition and funding:** Our aim is to participate any of the available EU, UNESCO or OECD recognition and funding schemes for our observatories. This enables a higher level of validation, recognition, and excess to base funding for a permanent data program.

### Who we are?

**Reprex** is a for-profit company registered in the Netherlands in 2020. Our diverse team is partly made of the freelancer network of the CEEMID project in 2014-2019 and new team members with a background in music and data. The co-founding shareholders are Daniel Antal, CFA and Réka Szentirmay. You can find more information about [our team](#) on our website. In 2020, we have successfully completed a granted work for the Slovak Arts Council, and validated our product/market fit in the prestigious [Yes!Deflt AI+Blockchain Validation Lab](#). You can **contact us** on various channels via [our website](#).