|the| technopolitan

The newsletter of Technopolis |group|
An international research business focusing on the evaluation and development of policy in the fields of research and innovation

October 2016 - N° 15

Big data and policy

Turning data into information and insights

'Big data' is opening up many exciting opportunities for evidence based policy making. While we are only starting to understand the implications of big data, policy makers around the world have started to experiment and implement its applications.

As a provider of policy advice, Technopolis Group uses big data — and supports its clients in their use of big data. Together with clients and partners we have looked at good practices and interesting new tools that can be applied in the policy fields that we are passionate about. Using the company's Innovation Fund, our colleagues have experimented with new data sources, analytical tools and visualisation techniques to add value to the traditional empirical methods normally used in the knowledge triangle.

This issue of |the| technopolitan gives you a flavour of the work we have carried out in the last two years. We hope it triggers new ideas on how big data can support your policy work.

Patries Boekholt, Group Managing Director
patries.boekholt@technopolis-group.com

Big data for evidence-informed policy: from experiments to implementation

'Big data' is about high-velocity, high-variety and high-volume data sources — such as sensor data, social media data, transaction data and satellite data — and about methods to analyse data and visualise the results. Using big data has become the norm for companies such as retailers, online media providers and insurance companies. Policy makers, however, have only started to explore the opportunities of using big data. This holds for explorative policy analysis that seeks to understand societal developments and transitions, and indeed for data collection and analysis for the design, implementation and evaluation of concrete policy interventions.

Over the last two years, Technopolis Group has had the privilege to analyse the use of big data by national and international policy makers. In a study for the European Commission (EC), Technopolis Group, the Oxford Internet Institute and the Centre for European Policy Studies (CEPS) interviewed thought-leaders, identified 58 state-of-the-art 'data4policy' initiatives, and organised an expert workshop. Our client was the EC's DG CONNECT. Several DGs were involved in scoping the study and in developing 10 use cases for the European Commission, in areas such as education, health and trade policy.

One of these use cases, on bee health, was developed into an online demonstrator. The concept should be applicable to many other policy areas that could benefit from real-time, continuous sensor monitoring combined with a comprehensive set of environmental and policy data that can be linked at the local, regional, national and EU level. Showing such combined data in a unified visualisation and analysis interface allows explorations and inferences that may not be apparent by looking at individual data sets.

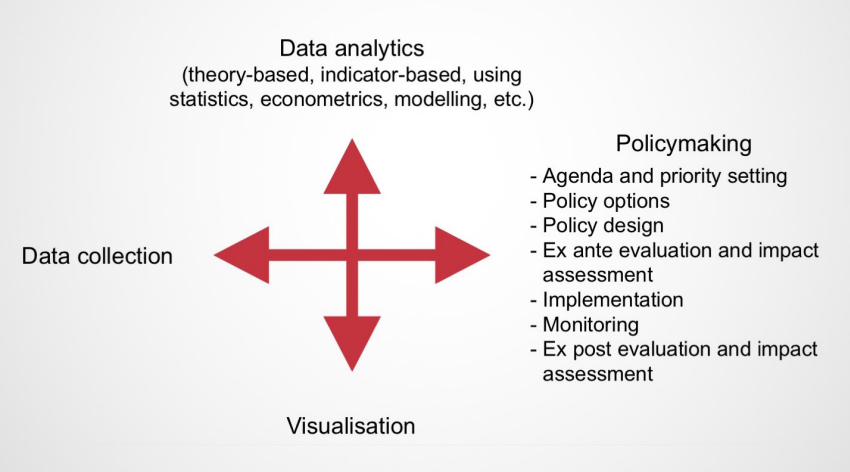
The results of the study are available at www.data4policy.eu

Trends and observations

In this introduction to this special issue of |the| technopolitan we share 10 observations made during the data4policy study and while experimenting with new data sources, data linking and tools for analysing unstructured data.

1. An overarching observation is that the use of big data by policy makers increases step-by-step, depending on the specific characteristics of policy areas, policy makers and the added value of new data sources. For instance, when official statistics based on large-scale surveys are lacking, or when the results are available too late, the incentives for using new data sources are substantial. An example is the use of mobile phone data (location data) to estimate population density, traffic flows and trade flows in developing countries.
2. Smart cities and local policy makers are well ahead of national and international policy makers, for several reasons: public-private collaboration and citizen engagement are easier to organise at a local level, sensor networks can be installed quickly at the city level, and local policy makers can clearly communicate the advantages of big data approaches, for instance when providing citizens with access to data about air quality, parking spaces and safety.

Data, analytics, visualization and the policy process



Source: Technopolis Group, Oxford Internet Institute and CEPS

3. National and, to a lesser extent, international policy makers are starting to use big data in areas such as transport, labour markets, and economic and environmental policy. Examples are the use of vehicle detection sensors to monitor traffic, text mining of vacancy databases to assess skills shortages, web scraping online retail websites to monitor inflation, and using sensor data to monitor water quality.
4. New data sources, such as social media data and other types of web data and unstructured data, are used in many research and innovation projects (such as Horizon 2020) but the actual implementation by policy makers is fairly limited. This reflects a normal process of developing and testing new data sources, and addressing barriers related to the regulation of data ownership, protecting privacy, the skills of data scientists and users, organisational routines, risk-aversion and cost, when implementing and scaling up these new data sources.
5. A similar evolution can be observed for the use of advanced analytical tools such as machine learning, profiling and predictive analytics such as 'nowcasting' (providing continuous forecasts of the near- and medium-term future). These tools are developed and used in research projects, while the emphasis in applied policy analysis is still on descriptive statistics.
6. Along the same lines, big data is used mostly during the first half of the policy cycle (e.g. foresight, problem analysis and policy design), while only slowly we see applications emerging in interim and ex post evaluation and impact assessment (an example of so called altmetrics is web scraping to assess the visibility of research institutes).
7. There is a certain hype element to big data but, if we unpack the concept and take an evolutionary perspective, we can see a risk of underestimating its impact. The possibilities for using big data for policy will increase as new data sources become more widely available (with some data being generated automatically, as a by-product of other processes), data linking becomes more feasible (benefiting from open data and using unique identifiers for regions, industries, technologies, etc.), analytical tools are improved (in functionalities and user-friendliness) and online visualisation tools become more intuitive.
8. Important challenges for using big data in the policy process, as opposed to using big data in companies, are the requirements for transparency, accuracy and accountability: in short, no 'black boxing'. For instance, algorithms must be understood by policy makers and the main elements communicated to politicians and citizens. Moreover, when data sources such as social media data are biased towards certain age groups and income levels, additional data sources are needed to ensure that all voices are included in the policy process.
9. Another challenge is to find a new balance, between explorative research approaches using the abundance of data on topics such as technology, innovation and economic trends ('ask the data') and carefully selecting the indicators and data that are most relevant for analysing such trends and assessing the impact of policy interventions ('start with the intervention logic'). To rely too much on readily available data brings the risk of using irrelevant data. For instance, popular studies based on Twitter data and Google search entries, turned out to be far less rigorous and relevant than was expected at the time of their publishing. In short: there are lies, statistics and big data.
10. One of the opportunities of big data is dual learning: launching policy experiments and setting up a data collection and data analysis strategy to monitor and assess how policy interventions change the behaviour of actors. We should acknowledge the limitations of established data sources and collect both types of data. Moreover, stakeholder engagement remains crucial (or becomes even more crucial) for interpreting data, creating insights and improving policy interventions.

For more information, please contact
martijn.poel@technopolis-group.com

Exploring business-university collaboration with 'Name Entity Recognition'

Technopolis Group has used machine-reading algorithms to classify and analyse text in a huge dataset — with the aim of illustrating the landscape of research collaborations between universities and companies in the UK.

As part of the 2015 Dowling Review of research collaboration between universities and the private sector, the Royal Academy of Engineering (RAEng) explored different ways to illustrate these collaborations in the UK. Technopolis Group was commissioned to build on prior work — for the RAEng and the Engineering and Physical Sciences Research Council (ESPRC) — which tested a new approach to obtaining company mentions from unstructured text information.

Our source of information was a database of 6,642 impact case studies submitted in the UK's latest Research Excellence Framework (REF) exercise, in which universities describe the impact of their research and their different networks and collaborations.

The information contained in these impact case studies amounts to 10,200 full pages of written text. This called for the use of automated machine-reading techniques. We used the latest version of the Name Entity Recognition (NER) classifier published by the Stanford Natural Language Processing Group.

Tagging and classifying

NER algorithms take text and tag it with specific metadata, locating and classifying expressions into different sets of predefined categories or entities. Some of these algorithms are pre-trained with large volumes of textual data and use advanced techniques to detect names in the text and classify them by type of entity. Using a 3-class model for English language, the NER classifier tags organisations, locations and personal names present in a text. As is usually the case when using these types of 'big data' techniques, some old fashioned manual cleaning was also put in place.

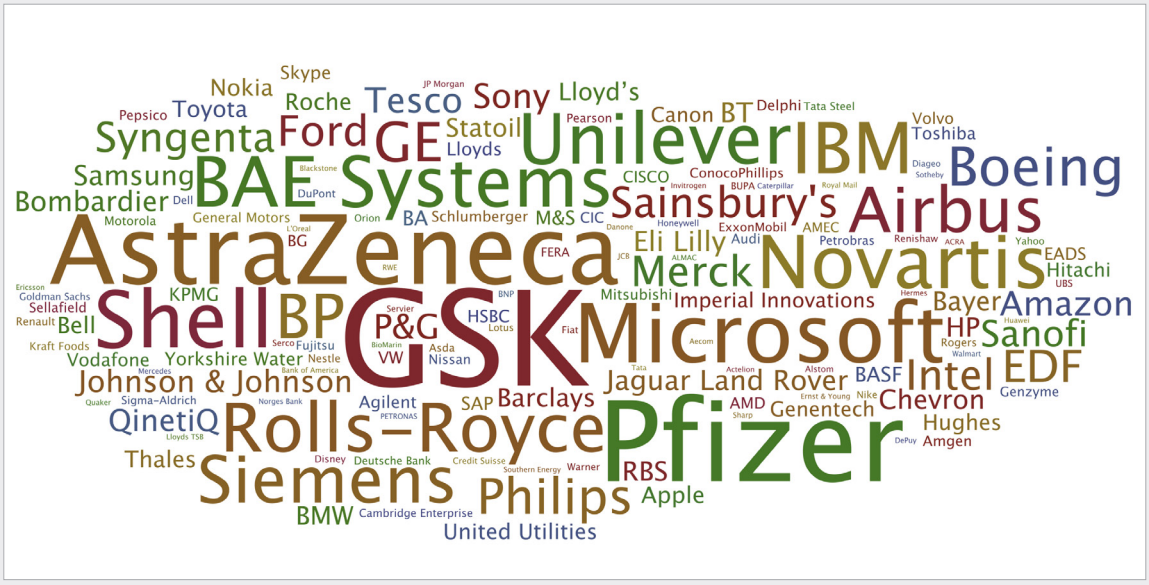
The final result is a simple but quite powerful set of word clouds identifying the companies that interact the most with universities — or are mentioned the most by universities — in the UK.

For more information, please contact

cristina.rosemberg@technopolis-group.com

xavier.potau@technopolis-group.com

Companies mentioned in 5 or more impact case studies (175 companies)



Using Wikipedia metadata to monitor Responsible Research & Innovation trends

Wikipedia is the largest and most popular general reference work on the internet and is ranked among the 10 most popular websites worldwide. A team of Technopolitans recently used it as a new data source from which to extract trends in biotechnology, environmental technology, medical technology and nanotechnology.

For many web users, Wikipedia is not just a source of information. Through its open editorial processes, it is also a space for debates on a range of topics – including science, technologies and innovation (STI). This makes Wikipedia an important driver of what innovation scholars call the "social shaping of technologies".

As a side effect, the trail of data generated by edit histories and discussions on Wikipedia can be exploited for 'big data' approaches. Using advanced data analytics tools, they can be used to derive quantitative indicators on debates about technologies and to approach questions in the area of innovation or technology policy such as:

- How extensively are articles on technologies worked on and discussed, i.e. how important are these topics for the Wikipedia community (as a proxy for society)?
- When do debates around technologies take off on Wikipedia and how does that correlate with technological trajectories?

'Wikipedia edits' is therefore a complementary indicator to social media statistics and newspaper articles, as Wikipedia data is generated bottom-up (like social media) but — due to its higher transaction cost — is a better indicator of the 'depth' of the concern (like newspapers), rather than just the breadth.

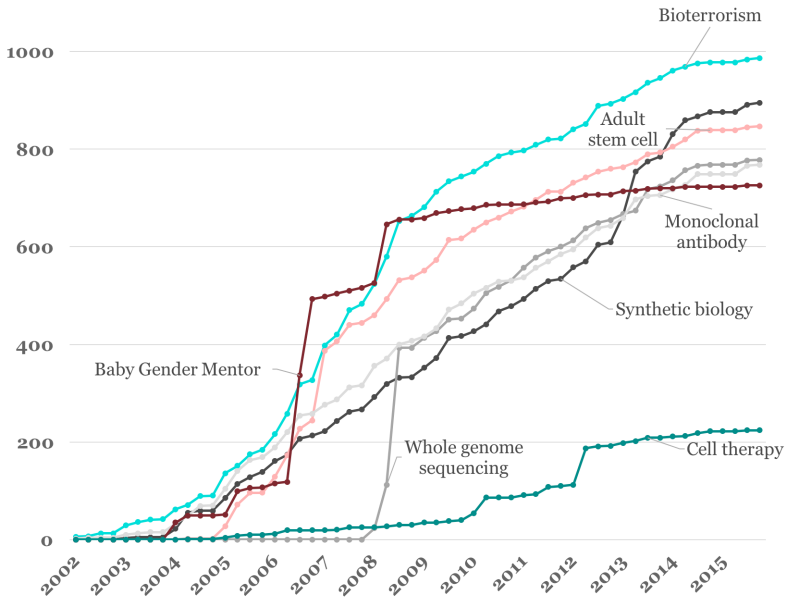
Trending topics, ethical debate

A team of Technopolitans recently tapped into this new data source. Given the importance of Responsible Research & Innovation (RRI), we looked into Wikipedia debates about technologies with clear ethical and normative aspects. Monitoring whether particular technologies and ethical concerns are 'trending topics' can be useful for policy makers in order to engage stakeholders and address concerns. We extracted data from Wikipedia pages on biotechnology, environmental technology, medical technology and nanotechnology.

The graph shows an analysis of article edits in Wikipedia for selected biotechnologies. The trends show which topics are discussed most intensively, and which are largely driven by piecemeal contributions from experts (e.g. the steady growing trend for 'monoclonal antibodies') or a sudden burst of public concern with limited follow-up. A good example of the latter is the 'Baby Gender Mentor', a controversial new product for very early gender screening of foetuses that was discussed on US television in late 2006.

For more information, please contact
florian.berger@technopolis-group.com

Number of Wikipedia article edits for selected Biotech pages (cumulative)



Using big data for sustainable development

The quantity, diversity and accessibility of digital information has expanded hugely thanks to innovations in technology (for usage, storage and analysis) and increasing affordability of digital devices. In developing and emerging countries, access to ICT and mobile technologies has expanded rapidly in recent years — with implications for 'big data'.

Searching for smart policies or new business strategies, public and private actors in these countries increasingly recognise the benefits of using these new sources of intelligence and evidence. Many companies are interested in using 'big data' analytics in emerging markets — with their budding middle class — to improve their marketing strategies and to increase the efficiency of their supply chains. For the public sector, big data can inform early-warning systems, real-time trend analysis, needs assessment, and monitoring and evaluation. Big data applications have already shown game-changing impacts in sectors such as humanitarian assistance, epidemics tracking, crop forecasting, traffic management, mobile banking and tax collection systems.

Although the digital divide is closing fast, privacy issues, inequality of digital access and limited generation of digital data in remote or rural areas remain a challenge, resulting in insufficient data and often inconclusive analyses for the public sector. The World Bank has made steps toward collecting digital data for emerging markets and making it widely available. Since 2009, the UN's Global Pulse initiative serves as an innovation lab to facilitate the adoption of big data tools and technologies. One of its pilot projects confirmed the potential of social networks, such as Twitter, to complement traditional household surveys when analysing public perception, in this case related to gas policy reform in Salvador.

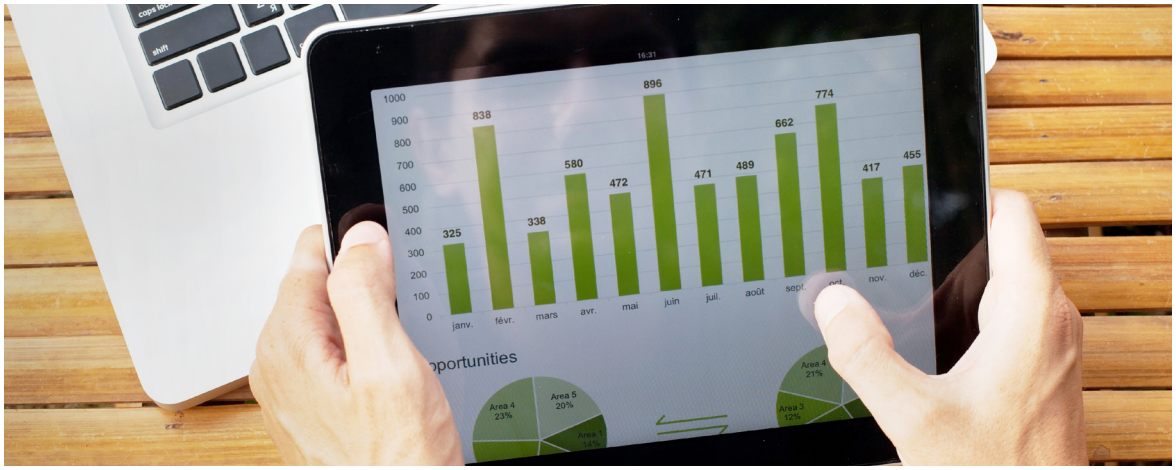
Big data opportunities

In our own work in emerging countries, big data is increasingly relevant both as a tool for studies, strategy and evaluation, and as a policy issue for clients including international organisations, development banks and development agencies. Technopolis Group experts contributed to the UN Conference on 'Big Data and the 2030 Agenda for Sustainable Development: Achieving Development Goals in the Asia and the Pacific Region', held in December 2016 in Bangkok, Thailand.

Big data already adds value to our studies and advice for clients working in emerging and developing countries. Examples include:

- **Higher education and entrepreneurship policies** that address skills needs for the use of big data and the interpretation of results
- Big data policies for **life sciences and health**, including management and use of biobanks and patient information databases and the use of social media data to analyse the effectiveness of policy interventions
- **Green growth** policy monitoring using geo-observational data
- **Scientific collaboration and research capacity building,** including 'citizen science', eg data collection via citizen-owned sensors and mobile phone apps
- **Research and innovation infrastructures** for data analytics, eg access to open data, High Performance Computing, data analytics software and visualisation tools
- **Data linking** to assess the interaction between higher education, research and innovation, economic development and labour markets in specific countries and regions.

For more information, please contact
francie.sadeski@technopolis-group.com
matthias.ploeg@technopolis-group.com



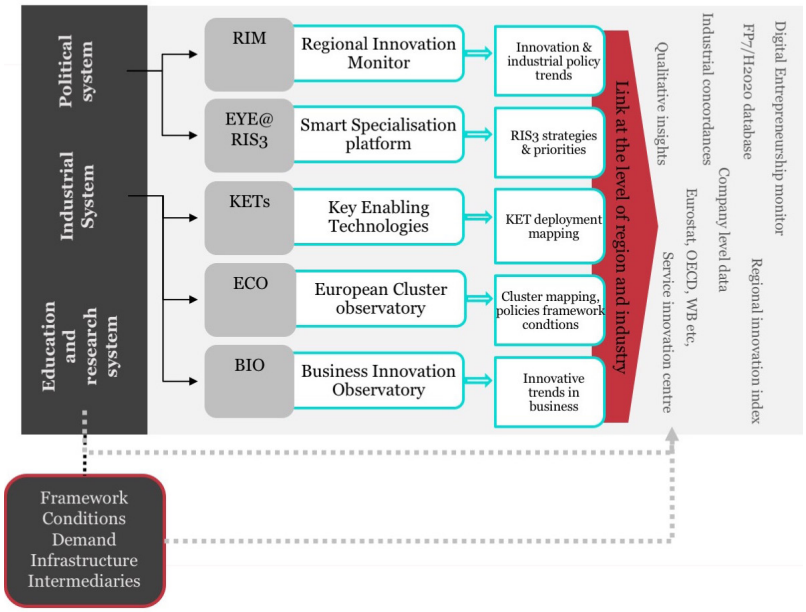
The next step in data linking

Integrating existing science and innovation databases to enrich our knowledge base for policy support.

In the field of Science, Technology and Innovation (STI), there is a tradition of large-scale data collection. The EU launched its first large-scale monitoring initiatives — the INNO Policy Trendchart and European Innovation Scoreboard — in 1999 and 2000 respectively, followed by several other STI-focused monitoring platforms, observatories and projects. This has resulted in huge datasets, collected through networks, surveys and statistical offices. Increasingly, this data is available in open data formats, via open data portals.

Is this big data? Probably not, but it is certainly a lot of data, especially if we link different types of datasets and use them to explore the interplay of large numbers of variables.

Example of Integrated database



Multiplying mess

Given the 'messiness' of individual STI datasets (missing data, etc.), one of the main challenges of such data linking is avoiding a 'messiness multiplier' when linking datasets collected for different purposes using different designs (e.g. different perspectives, definitions, data collection methods, reliability levels, etc.).

For example, consider the contrasts between business and policy-maker centred designs (e.g. Business Innovation Observatory vs Regional Innovation Monitor), or between thematically focused (i.e. policy, industry or technology specific), STI-system encompassing (Key Enabling Technologies Observatory, European Cluster Observatory or European Innovation Scoreboard) and value-chain inspired designs (e.g. the up-coming RE-Confirm).

So it is natural to question whether an integrated monitoring system would be of greater value than the sum of its parts — and its ability to support policy design.

Technopolis Group recently used data linking in a comparative analysis of the national 'Smart Specialisation Strategies' in Central Europe. The study identified seven common 'technology priority areas' using data from five different sources: the Smart Specialisation Platform (EYE@RIS3), FP7/H2020 databases, the European Cluster Observatory, the Key Enabling Technologies Observatory and OECD trade data. This selection was then verified based on analysis of R&D capacities (upstream), clustering and patenting (midstream) and business or economic activity in the respective technology areas (downstream).

The next step

A takeaway from the project is that the public sector needs support in anticipating new developments. Traditional STI databases are useful but have limitations when it comes to informing policy at a high level of granularity (e.g. in terms of industry segments/niches) or capturing dynamics on the ground.

An integrated monitoring system would make such work more efficient and enable more data to be consulted in one go ('integrated dashboards') — e.g. to test hypotheses and identify information gaps. But to make the next step, and truly enrich our knowledge base, we need to turn the traditional STI indicators collected by these observatories and other projects into 'real' big data: In other words, data linking that uses reliable STI concepts and structures for markets, industries and regions, but also acknowledges the strengths of using web scraping, social media, and Wikipedia data to identify emerging technologies, new application areas and new markets.

For further information, contact:
paresa.markianidou@technopolis-group.com

Technopolis Group news

News from Colombia

In August, senior leaders from Technopolis Group were invited to participate in the international conference "Colombia + Innovative: building linkages for accelerating innovation." The event was organised by The Office of The President of Colombia, the National Planning Department and the Swiss Embassy in Colombia. It was opened by the President of the Swiss Confederation, Johann N. Schneider-Ammann, and by high representatives of the Colombian government.

Technopolis Group Chairman, Dr Erik Arnold delivered the keynote speech to the conference on "Systemic Innovation Policy", and Dr Patricia Boekholt, Technopolis Group MD, was a member of the expert panel, which discussed "Desirable and undesirable patterns of innovation systems." During their visit to Bogotá, Erik and Patries, along with

Juan Carlos Salazar, Director, Technopolis Group Colombia, also participated in a workshop with members of the National Technical Committee of Innovation, and attended meetings at Los Andes University, the Bank of Development of Latin America and the Delegation of the European Union in Colombia.

Follow Technopolis Group

Information

Subscription

To receive |the| technopolitan electronically, please sign up for our newsletter via our web site.

Contact

www.technopolis-group.com

Amsterdam |NL|
Technopolis BV
Spuistraat 283
1012 VR Amsterdam
T +31 20 535 2244
info.nl@technopolis-group.com

Bogotá |CO|
Technopolis Group Colombia SAS
Cra 15A Bis No 45-65
Bogotá
T +57 1 7022280
info.co@technopolis-group.com

Brighton |UK|
Technopolis Ltd
3 Pavilion Buildings
Brighton BN1 1EE
T +44 1273 204320
info@technopolis-group.com

Brussels |BE|
Technopolis Belgium
Avenue de Tervuren 188A
B-1150 Brussels
T +32 2 737 74 40
info.be@technopolis-group.com

Frankfurt/Main |DE|
Technopolis Deutschland GmbH
Große Seetraße 26
DE-60486 Frankfurt am Main
T +49 69 348 7679 80
info.de@technopolis-group.com

Paris |FR|
Technopolis France S.A.R.L.
88 rue La Fayette
75009 Paris
T +33 1 49 49 09 20
info.fr@technopolis-group.com

Stockholm |SE|
Faugert & Co Utvärdering AB
Skeppargatan 27
114 52 Stockholm
T +46 8 55 11 81 00
info@faugert.se

Tallinn |EE|
Technopolis Group Estonia
Narva mnt 5
Tallinn 10117
T +372 644 0435
info.ee@technopolis-group.com

Vienna |AT|
Technopolis Austria GmbH
Rudolfsplatz 12/11
A-1010, Vienna
T +43 1 503 9592 - 13/14/17
info.at@technopolis-group.com