Summary

Our country needs more open data, of all kinds. This includes open government data, and access to and sharing of sensitive information. France has been a pioneer of open data and open source policy in Europe since 2013, but our lead is precarious. Our policy is currently caught up in the wrong debate – "for or against open data" – and the outcome is a shift in policy objectives, whereas we should be changing the way we do things. Our inaction has resulted in lost opportunities for our society and our economy, while the potential for knowledge development and innovation is huge.

This report sets out some ambitious reforms, in particular if we want to take part in the transformations under way in Europe, as well as many realistic measures that do not require a "new dawn" and can be achieved within a year and have a significant impact on the future. If the government does not promote this policy, France will miss out on a major opportunity to build trust in government, improve the efficacy of public policy, and develop knowledge and innovation across the economy. An urgent need for such action has now emerged from the health and economic crisis and the findings of the "Great National Debate" in France.

The Public Interest Entrepreneurs programme in France, which involves hiring talented individuals in government departments to solve targeted problems, has shown that even a modest level of commitment can deliver substantial results when open data and open source technology are used. France's 2016 Digital Republic Act has also shown that the country can conduct major legislative reform while continuing to rank as a daring and innovative nation. This report lays out the choice our country is facing today.

It's worth repeating: an open data policy is good for everyone

The benefits of open data and open source still need to be explained and proven, even in the wake of the COVID-19 crisis, which has shone a spotlight on the importance of data in public policy.

Many people, especially within the government, do not understand this policy or its objectives: "we haven't been told why we need to make data open". Neither do they understand the impact of reusing data and source code (this impact has not been measured). At best, open data is perceived as something they have to do; at worst, they have the impression that it is someone else's job: one government department says they do not handle any data "because they are not a statistics department".

Open data and open source are not just "tech" issues. They are first and foremost political, democratic, scientific and economic issues. The beginnings of an open data policy can be found in the 1978 Act, which laid the foundations for the right to access government documents, which include open data and open source.

From a scientific standpoint, data disseminates knowledge; researchers share the data and source code they use in what is known as "open science". If teams of researchers were better at sharing their research, the treatment of COVID-19 patients would have certainly been faster and more effective during the crisis. More broadly, in all fields of knowledge, data is also the building block of artificial intelligence (AI). We are just starting to realise the potential of AI.

Economically speaking, data drives innovation. When the French property price database (DVF) was made open data in 2019, a host of innovative services and businesses were created in France and other countries to analyse house prices. A European Commission study in 2019 valued open data in France at €28 billion. The use of open source software also drives growth and should be the focus of an internal innovation policy for government action. France's National Cybersecurity Agency (ANSSI) is a trailblazer in the field, demonstrating that sharing source code does not make an organisation's information systems vulnerable.

In the democratic arena, data improves the public service by connecting government with users. The people of Taiwan have a platform they can access to manage the information they share with the government. They can link their information with a government service of their choice and update the information they provide all government services in one click. Why not have a similar service in France?

In politics, data is a good way of restoring trust in government. The challenge is huge: our public consultation revealed that people had a lot of mistrust in the data used by the government. The COVID-19 crisis has raised the nation's awareness of data. What we have witnessed during the crisis – problems in providing live statistics about the epidemic, different interpretations of epidemic curves, and uncertainty surrounding the assumptions of epidemiological modelling – put us all on a steep learning curve. As a result, we came to understand that all data is constructed and needs to be explained. Making data open enhances the public debate. Transparency is the best way to dispel mistrust and conspiracy theories.

The crisis also showed that data-driven government increasingly requires accessing data held by the private sector, without which it would be impossible to track what is going on in the country: take, for example, Google's community mobility reports, Orange's population mobility data and Crédit Mutuel's reports on bank account use.

Open data is also a way of properly measuring public policy. Producing data is not just about making good old-fashioned audits more reliable or conducting live audits; it is also about tracking how public money is spent. This is precisely the aim of France's Government Action Barometer. Also, in order to thoroughly evaluate any public policy, we have to share public policy data securely between government departments and with researchers. France's government statistics agency is no longer restricted to the realm of national accounts; they now act as a network of statistics departments tasked with carefully assessing the programmes of each of the ministries where they work.

Open data is very far from realising its full potential. There is no truth in saying that open data has delivered disappointing results since the policy was started in 2016 and it should be ended. Our open data policy is largely incomplete and even unsatisfactory in some respects of how it was implemented. Government agencies do not always understand when open data can be reused. This is another instance where we need to change how we do things, not our policy objectives.

Open data: striking the right balance

We need to strike a balance between open data and protection. The EU seems to have found a balance in the General Data Protection Regulation (GDPR) for personal data, but France has not consistently achieved this in its legislative framework or current practices, which do not harness all the flexibility provided for in the Regulation.

The EU's personal data protection rules should be effective in practice and should not be systematically over-interpreted.

Security should not be used as an excuse. Some government agencies are using security as a pretext not to make their data open and are interpreting security in their own interest. The taskforce recommends that ANSSI play a greater role in the open data policy in order to provide expertise in information systems security and argue that open source software is more secure than closed source software.

Not providing open data can be "political" because of a fear that it might be questioned or misused: the taskforce recommends that the Institut Pasteur's forecast models for COVID-19 and some delinquency statistics be made open data. Data does not have a political purpose, but the way it is reused does. When there is hardly any open data, the focus turns to whatever figures are available. Open data promotes a debate based on facts. Government departments should be able to make their voice heard in the debate, providing their expertise on how to interpret data. Other voices should also be heard and contribute to the debate.

Open data should be more widely available. The principle of data being open by default, which was enacted in 2016, means that governments release data, as opposed to individual citizens requesting access to information,

i.e. as provided for under the 1978 Act. Many government departments are not meeting their obligations. We therefore need to make the law more effective by increasing the powers of France's Commission for Access to Administrative Documents (CADA). In 80% of instances[1], the government does not even answer requests and the average response time from CADA was 176 days in 2019.

Lastly, we need to remove any unnecessary barriers to open data and open source that receive public funding by looking at the fees that are still being charged by government departments and the intellectual property system for public servants, who should only be paid for their work. In the case of the private sector working for the government, there should be a guarantee that government data is made available to them as it will have an economic impact on data producers, such as registrars of commercial courts.

Improvements needed to quality and accessibility

Open data maturity needs to be improved without forsaking open data's basic principles, i.e. open by default and able to be freely reused. There is no point coming up with selection criteria to determine which datasets should be made open first because there is no way of knowing in advance how open data and open source will be reused. We also cannot fully predict why data should be made open. Open data is also about listening to the needs of civil society, rather than waiting for them to become points of contention. Similarly, licences restricting reuse are not advisable.

Open data has to enter a new era and aim for improved data quality and reliability. Some avenues to explore are improving documentation (which is often incomplete), formulating interoperability standards and using more consistent metadata. Open government data should be expanded to include new benchmark datasets to improve quality and availability. Quality is critical to the development of artificial intelligence, which can be supplied government data or sensitive information, but the data needs to be high-volume and high-quality.

Listening to reusers will deliver quality. At the moment, there is rarely any connection between data producers and reusers. The government department that produces the most reused dataset available on data.gouv.fr (DVF) does not regularly participate in the work of the community of reusers, who could provide some valuable input on the quality of the data-producing department. One exception worth highlighting is the relationship between ministerial statistics departments and researchers. These statistics departments provide "user feedback" and contribute to research, but they are not actual ministerial departments.

Above and beyond the inherent quality of the data, data dissemination also needs to be high-quality in order for information to circulate: if data-producing departments use a system of quality seals, it is easier for the data to be reused. Data infrastructure should come up with a response by providing secure interoperability and appropriate distribution services. Their governance models should include reusers. Government investment, starting with programmes funded under the recovery plan, should incorporate data distribution.

Open data should also be made more accessible and easier to find. Data should be featured in data catalogues that are easy to find and explore in order to clearly identify data producers. The data.gouv.fr website should be upgraded to make information easier to explore and access. Creating APIs to access databases, such as the SIRENE database managed by France's National Institute of Statistics and Economic Studies (INSEE), can discourage users and should be used proportionately – even though it allows closer monitoring/analytics of reusers and can provide some user management functions.

For government departments using open source code and open source software, we need to bring them together in a community and provide them with more support. Open source software is not an ideological movement completely separate from the business of government departments; it is not just a matter of who is using LibreOffice. It provides an opportunity to finally share and pool data in the public sector so two departments are not wasting time working on the same policy issue without knowing what the other department is doing. Open source software also has mutual benefits for the government and the economy as they work together to develop programmes in the public interest. It is a way of attracting top tech talent to government jobs. The taskforce recommends setting up an Open Source Program Office (OSPO) – a high-profile, long-term initiative – within the Interministerial Directorate for Digital Technology (DINUM). This would be a starting point to addressing the challenges of open data and open source.

[1] Sample of 98 requests submitted by the L'Ouvre-Boîte organisation between 2017 and 2020.

Government departments should share data to improve the effectiveness of government action

Some data cannot be open to everyone. This is when you start sharing and accessing data. Sharing is when a user has a physical copy of the data on their server and accessing it is when they can only use the data by accessing a data producer's server and they cannot keep a physical copy of it.

Unfortunately, many people have a reductive view of open data and do not even consider the possibility of sharing some data in a restricted, secure environment. This is often because they do not trust data reuse. One way of getting around this issue would be for the government to promote the policy throughout its highest ranks.

There is very little sharing of data between government departments – and this is shocking. Some departments even go as far as to re-enter data that is available in another department of the same ministry. Open data is sometimes the only way a department can find out about another department's data and access it. This is another argument in favour of free and open data – where possible.

When data can be shared, the procedure for sharing it can be too difficult: this is the case when departments want to use a person's National Registration Number (NIR) to cross-reference data in two different databases. This secure procedure was made easier under France's 2016 Digital Republic Act, but four years later, it is still not available. There is currently no way of knowing the employment status of the large numbers of income support recipients recorded this year.

Sharing information is also restricted between the central government and local and regional authorities, despite some instances of cooperation on regional data platforms. But local and regional authorities normally do not provide any access to data about how national programmes are implemented under their watch, even though these programmes are fully funded by the central government, e.g. economic development programmes. With the government's response to the COVID-19 outbreak, there was a strong need for information from local and regional authorities about epidemiological trends in each region. This was used to monitor local lockdown measures and curfews.

Secure access to sensitive information should boost our AI independence

Securely accessing data is a way of analysing data that is stored on a data owner's server, providing maximum protection of sensitive personal information. Secure access allows all the value of the data to be extracted, while guaranteeing strategic independence for researchers. Secure access is particularly useful for researchers as they need to use data that is not anonymised, but personally identifiable or pseudonymised (a lower level of anonymisation, preventing re-identification without the use of additional information).

Secure access applications are becoming more common with the emergence of sector-specific or inter-sector data hubs and governance tools, thereby helping to remove data silos. Some examples include the Health Data Hub, the Agdatahub and the Secure Data Access Centre (CASD), which was initially designed as an offshoot of INSEE for researchers. It is good to see hubs developing in different sectors, but these should not be siloed and prevent inter-connection in the future.

There are not enough datasets in France to train algorithms for artificial intelligence at the moment. As a result, French start-ups have to look for datasets in other countries in order to develop the tools and services we use on a daily basis. **We should do everything we can to guarantee our strategic independence in AI or our lives will soon be run by algorithms trained from data which does not reflect our values or our principles.**

There are still too many restrictions around researchers accessing data. We have made some good progress, but France is lagging behind international standards. We need to improve our handling of researchers' requests for information and include data officers and ministerial statistics departments in the process. Research is essential in evaluating government action. One French researcher has been trying to research posted workers in France and has still not received the information requested from the government more than two years ago, even though she has been granted approval from the committee for statistical confidentiality. She had no trouble getting the information from Belgium, Luxembourg or Portugal.

Governments now need to use private-sector data on a large scale

"Data in the general interest" is a term used to define instances where private-sector data can be made available "in the general interest". Since there is no actual definition, there are two scenarios which require different responses from the government: first, the government uses data produced by the private sector (business-to-government, or B2G), e.g. throughout the COVID-19 crisis, the government used data from Orange and credit card companies to monitor activity in the country during lockdown. Second, the private sector shares data in a particular sector, for example (business-to-business, or B2B). Private-sector data provided as open data because they are operating a public service is not necessarily data in the general interest, but open government data. One example is the release of public transport data (timetables, fares and payments, and trip planning).

If you look at France's history, the government has often favoured national producers of information over private producers in order to guarantee autonomy in decision-making. One example is the National Institute of Geographic and Forest Information (IGN), whose origins can be traced back to military requirements in the 17th century. Another example is INSEE, which was founded in 1946 and had a "monopoly" to produce economic data. France is different to Germany and the United States in this respect. Data used to be conceived as political in France and did not have other uses[2].

It is already mandatory for the private sector to provide a range of information to the government or to other private-sector organisations under regulatory requirements, for example. For data in the general interest, new developments include using large datasets on a large scale, including big data, and using data for purposes other than statistical surveys.

It is not so much a question of the government's legitimate role, but of legal certainty. This data can only be made available in accordance with the freedom to conduct a business and the property rights that may apply to the data, with a guarantee of transparency if the data is reused by the government. The legal frameworks governing B2G and B2B need to be clarified in order to reassure the private sector about the process.

We now have to ramp up the sharing of data in the general interest, particularly within the scope of the EU (Data Governance Act, Data Act, Digital Services Act). This is a window of opportunity we cannot afford to miss. The taskforce believes that the work carried out to identify sectors which should share their data (one of the remits of a taskforce in 2016) should be the task of a formal agency and start operating as such. We need a cross-cutting organisation whose role would be to promote the need for more sharing of private-sector data. CADA succeeded in getting the government to listen to the needs of civil society. Now we have to work out which government departments need private-sector data.

Instead of establishing a single, holistic legal framework, which would not be an option in the short term because the system is not mature enough, we should manage these needs with a "project team" organisation. The relevant players would come together to work out how to go about sharing the data and under which governance model, with oversight by the government. Transparency and accountability should be the guiding principles. The issue of data being used by the government cannot be resolved between the government and the private sector alone because the data being shared is data about individuals. Representatives of civil society also have to be involved in the process.

Similarly, citizen-led data portability initiatives for data in the general interest should be encouraged in order to allow citizens to have better control over their personal data and how they choose to make it available, including in the general interest.

Lastly, the development of secure data sharing infrastructure, with built-in tools to manage data rights and rules, is essential to building the private sector's trust in the new culture of data-driven collaboration and value creation.

The necessary task of promoting the policy

What will happen after this report? As things stand today, we are afraid that our recommendations will not be publicised or followed. We have therefore put forward several recommendations to help with policy implementation at ground level.

[2] Pierre Rosanvallon, *L'État en France de 1789 à nos jours*, 1993.

First, the policy should be promoted by the government throughout its ranks: the issues raised in the report should be addressed on a more regular basis. We should not have to wait for a special taskforce to be formed. We need a government priority, a chief data officer and a higher profile and more support for ministerial data officers. There is no magic formula. Data should be an issue for the government; the prime minister should be at the forefront, chairing an interministerial committee. Data should also be promoted by the highest ranks of government: this is the responsibility of the Interministerial Directorate for Digital Technology (DINUM), tasked with implementation, and the Interministerial Directorate for Government Transformation (DITP), which monitors the policy.

Second, data is a new responsibility for public services and needs human and financial resources to match: France's Data Protection Authority (CNIL) and CADA should be able to respond to the growing number of complex requests for information, DINUM should be able to provide support to the government, and statistics departments should be available and responsive. This recommendation is not a drain on government resources because, as we know, an open data policy significantly boosts productivity if it is adequately funded.

Third, the recruitment policy for digital talent should be adapted. There has been some progress made in digital skills and talent management to attract and retain individuals with specialist skills, but not enough is being done to draw top talent in open data and open source technology.

Who knew that three of the previous 18 Debian project leaders were French? Debian is a major US developer of free and open source software. France has to support its tech talent just like it supports its high-level athletes and attract high-quality individuals to the government sector.

Fourth, the public service needs to be instilled with a culture of open data and open source. This was particularly highlighted during the public consultation process. The challenge lies in changing the culture of existing government employees, not just hiring new ones: too many senior public servants in leadership positions are afraid of open data, often out of ignorance. We cannot wait for future generations of tech- and data-savvy individuals to take up these posts. There is hardly any incentive at the moment for public servants to upskill and contribute to their department's data-driven transformation.

The taskforce has detailed several use cases in its report. These were chosen because they reflect some of our observations and our recommendations. We weighed up both sides of the issues with the relevant people in order to sufficiently establish the facts.

The taskforce would also like to note that the public consultation process between 8 October and 9 November 2020 provided a number of valuable contributions. A summary has been appended to the report. We have also highlighted the observations and recommendations which were formulated after meeting with a number of individuals during the process.

Recommendations

Cross-cutting recommendations

Recommendation No. 1: Start a public debate about trust in digital technology, outlining the basic principles of security and transparency for the government to address.

Recommendation No. 2: Engage with civil society, through citizens' consultations and the Open Government Partnership Forum, to determine which datasets and source code should be made open.

Recommendation No. 3: Conduct an assessment of the economic, social and scientific impact of making data open and sharing data and source code.

Policy promotion

Recommendation No. 4: Encourage the prime minister to promote the open data and open source policy. Add policy implementation and monitoring to the agenda of interministerial committee meetings chaired by the prime minister. Draft a circular letter outlining policy settings (governance, tasks and responsibilities in the government departments, interoperability, quality, legal guidance).

Implementation of open data and open source policy

Recommendation No. 5: Appoint a chief officer in charge of data, algorithms and source code (AGDAC) at DINUM with a reporting line to the prime minister. This full-time role would involve overseeing France's national open data and open source strategy with the support of ministerial officers in charge of data, algorithms and source code (AMDAC).

Recommendation No. 6: Steer and monitor the open data and open source policy at interministerial level (performance indicators, inclusion in impact assessments of government bills).

Recommendation No. 7: Encourage the government to play a more active part in digital commons.

Recommendation No. 8: Set up an Open Source Program Office (OSPO) or an open source software taskforce within TECH.GOUV, whose role is to help the government make public source code open and reuse it, identify opportunities for sharing source code, build ties with existing open source communities and support French tech talent.

Recommendation No. 9: Expand the powers of ministerial officers in charge of data, algorithms and source code (AMDAC) by:

- redefining their responsibilities in a standard job description
- providing ministerial officers with a mission letter signed by the relevant ministers after consulting with the directorates general and DINUM
- making sure that ministerial officers have enough resources to do their jobs
- systematically training ministerial officers and data protection officers together

Recommendation No. 10: Mandate the National Agency for Regional Cohesion (ANCT) to support regional authorities in publishing data and source code through programmes funded jointly by the central government and regional authorities.

Recommendation No. 11: Use open source and open data more extensively to promote French research in project assessments and funding.

Legislation and regulations

Recommendation No. 12: Amend the right to access government documents to make legislation more effective and allow the Commission for Access to Administrative Documents (CADA) to apply penalties for non-compliance with the provisions of France's code on relations between the public and the government (CRPA) related to the disclosure and publication of data and documents. Reduce the number of standard requests processed by CADA and streamline repeat requests made to CADA.

Recommendation No. 13: Assess the human resources requirements of the Data Protection Authority (CNIL) in order boost its advisory and support roles and monitor its increased resources with performance indicators based on user satisfaction (under the budget bill).

Recommendation No. 14: Appoint two new members to the boards of CNIL and CADA – one specialised in information systems security and the other in new data uses.

Recommendation No. 15: Involve the National Cybersecurity Agency (ANSSI) in the implementation of the open data and open source policy to make sure the policy does not affect information systems security:

- allow CADA and CNIL to refer a matter to ANSSI for its opinion to resolve any major information systems security issues
- allow the chief officer in charge of data, algorithms and source code (AGDAC) to request an audit from ANSSI of sensitive open source libraries and software

Recommendation No. 16: Check whether legislation allows all government data used by the private sector to be made open (i.e. regulated professions in the judiciary).

HR policy and cultural change

Recommendation No. 17: Formulate a more ambitious training policy in digital technology for the public service (mandatory tech training for senior public servants, seminars for senior public servants, training for public servants of all ranks, ministerial training programmes in addition to interministerial programmes, tech modules incorporated into all public service training courses).

Recommendation No. 18: Continue the work on digital skills and talent management and incorporate a tech component in initial public service training for digital professionals, building pathways for technical staff and offering permanent positions to contractors who hold skills that cannot be found in the ranks of existing public servants.

Recommendation No. 19: Diversify career pathways for INSEE "administrateurs" and officers in all government departments – not just ministerial statistics departments – and provide incentives to public servants who choose this career path.

Recommendation No. 20: Make a more compelling case for government employment targeting digital talent in short supply (mandatory remuneration benchmarking, reaching out to specialist higher education institutions).

Recommendation No. 21: Ramp up the Public Interest Entrepreneurs programme over the longer term.

Recommendation No. 22: Provide a training programme for elected officials about open data and open source in public policy.

Data quality

Recommendation No. 23: Develop a seal for data-producing departments to acknowledge their efforts, i.e. as part of the government data service.

Recommendation No. 24: Formulate and implement an interministerial interoperability and data quality policy (standardisation, FAIR data principles, metadata strategy, open data catalogues).

Recommendation No. 25: Encourage ecosystems to formulate a quality governance framework, appoint a quality officer and build communities of reusers with the active involvement of data producers.

Infrastructure, sharing and secure access

Recommendation No. 26: Channel investments under the recovery plan towards infrastructure to promote data dissemination (DINUM calls for proposals and sector-specific calls for proposals).

Recommendation No. 27: Encourage the creation of sector-specific or inter-sector data hubs under a targeted programme for each sector with guaranteed interoperability.

Recommendation No. 28: Set up a data sandbox to allow CNIL to gain exemptions from existing legislation to authorise the reuse of personal data in algorithms for artificial intelligence and store the data for longer periods than originally authorised when first collected.

Recommendation No. 29: Implement technical upgrades to use a file matching procedure based on "non-significant statistical code" for the purposes of government statistics and scientific and historical research.

Recommendation No. 30: Improve handling of researchers' requests and involve ministerial officers in charge of data, algorithms and source code (AMDAC) and ministerial statistics departments (maximum acceptable response time, appeal procedure, optional consultation with the committee for statistical confidentiality).

Data in the general interest

Recommendation No. 31: Take an incentivising and coordinated approach. Provide reasons for a coercive approach and conduct an assessment beforehand.

Recommendation No. 32: Secure the legal framework for voluntary sharing of data in the general interest relating to personal data (in a CNIL compliance guide) and the right to access and reuse private-sector data received by government departments.

Recommendation No. 33: Encourage citizen-led data portability initiatives for data in the general interest, i.e. by organising citizen engagement campaigns.

Use of private-sector data by the government (B2G)

Recommendation No. 34: Clarify the legal framework related to requisition to allow the government to access private-sector data in the urgent general interest.

Recommendation No. 35: Task the network of data officers (AGDAC/AMDAC) with the role of facilitating and mediating an arrangement for private-sector data to be accessed and used by the government (B2G), in tandem with the Directorate General for Enterprise (DGE).

Recommendation No. 36: Ensure the effectiveness of the provisions relating to data in the general interest in the Digital Republic Act which are encountering problems in the application phase for:

- data held by public service concessionaires and delegatees (standard clauses for government departments)
- private-sector data for statistical purposes – look into expanding Article 19 to include some data-based services

Sharing of data in the private sector (B2B)

Recommendation No. 37: Promote the sharing of private-sector data in the shared interest (B2B) on sector-specific strategy committees and in government calls for proposals (Invest for the Future Programme) and support initiatives by NGOs and the private sector.