



Building smart city data infrastructure: recommendations for the REPLICATE project

This activity has been developed under the REPLICATE project which is funded by the European Union's HORIZON 2020 Research and Innovation programme under grant agreement No. 691735.

Distributed by the Open Data Institute and Bristol City Council under a [Creative Commons Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/).



Context for this report

This paper has been produced by the Open Data Institute, at the request of Bristol City Council, to provide a set of recommendations for the REPLICATE project – an EU funded, integrated smart city project. It is intended to guide the project partners in their use of data and their approach to openness. While the recommendations are specific to the REPLICATE project, many will be applicable to other smart city projects.

About REPLICATE

[REPLICATE](#) is an EU funded (H2020) project in which 36 partners in three lead cities – Bristol, San Sebastian and Florence – are trialling integrated smart city solutions across three themes: energy, homes and mobility.

The project is building a variety of different services in these three cities, including; installing smart appliances and solar cells in homes, creating a platform for managing energy demand, installing electric car charging points and creating an on-demand electric minibus service, among many others. All these services will be joined up through a ‘smart city platform’ which will be used to create an integrated smart city solution.

The main aim of the trial is to create services which can be made sustainable, extensible and replicable in the lead cities and other cities around the world.

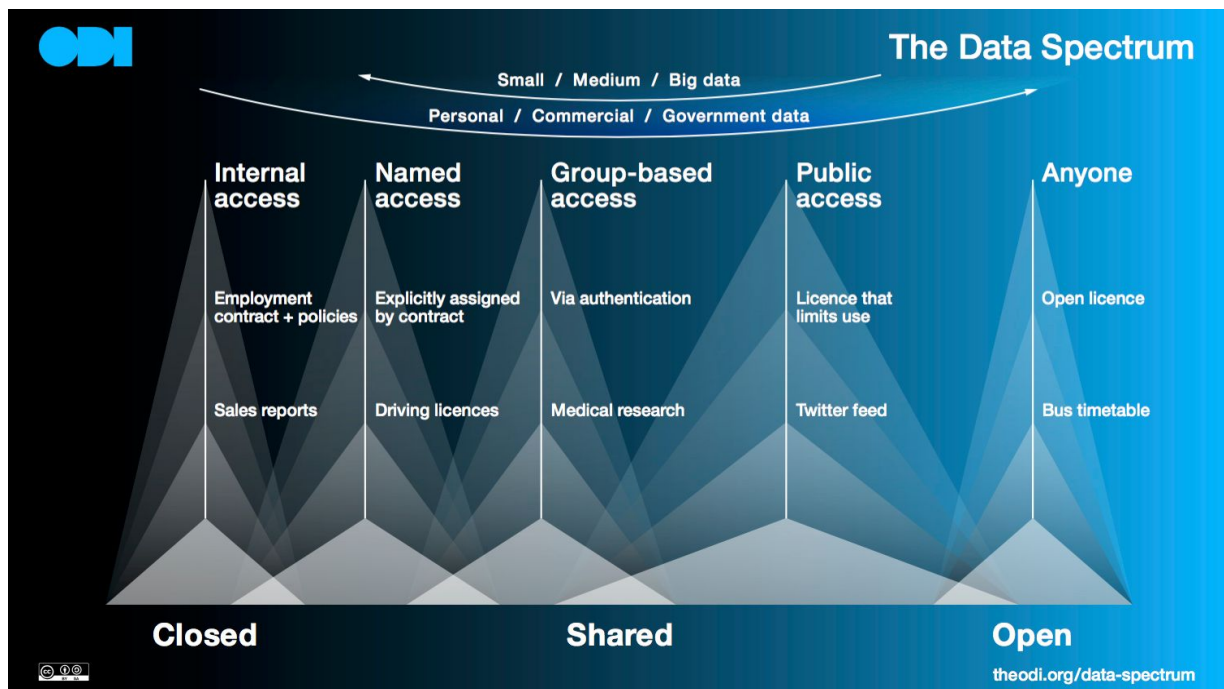
About the ODI

The [Open Data Institute](#) (ODI) helps governments and businesses around the world to get data to people who need it. It is independent, nonprofit and nonpartisan, founded in 2012 by Sir Tim Berners-Lee and Sir Nigel Shadbolt. From its headquarters in London and via its global network of startups, members and nodes, the ODI offers training, research and strategic advice for organisations looking to explore the possibilities of data.

Introduction: smart city data infrastructure

[Data is infrastructure](#). In a smart city, data is a key layer of infrastructure which helps connect new and existing physical, digital and social infrastructure to tackle urban challenges. By creating a strong data infrastructure, the REPLICATE project can ensure that the many components across its three themes can be effectively connected. By building this infrastructure according to best practices, it will also help ensure that the trial is replicable, sustainable and extensible beyond the lifetime of the project.

A data infrastructure consists not only of data assets but also the organisations that operate and maintain them and the guides describing how to use and manage the data. It includes technology, processes and organisation. The data assets in our data infrastructure may be in the shared or open parts of the [Data Spectrum](#). Trustworthy data infrastructure is sustainably funded and has oversight that provides direction to maximise data use and value by meeting the needs of society. The more open the data infrastructure, the more [value](#) it will create. It is crucial that we protect data that needs to be kept private, just as it is crucial that we openly publish data that should be open for everyone to use.



Context of the recommendations

The Open Data Institute was asked by Bristol City Council to work with the REPLICATE partners, primarily those involved in the Bristol pilot, to advise on how the project should handle closed, shared and open data.

To do this we have used the ODI's [principles for strengthening data infrastructure](#), our ['openness principles for organisations handling personal data'](#) and the [Data Spectrum](#). We will focus specifically on the data infrastructure principles: 'design for open' and 'respecting privacy', as well as some of the key challenges for REPLICATE that arose from a half-day workshop held in Bristol in April 2017 and several short interviews with partners.

ODI principles for strengthening data infrastructure

1. Design for open
2. Build with the web
3. Respect privacy
4. Benefit everyone
5. Think big but start small
6. Design to adapt
7. Encourage open innovation

ODI openness principles for organisations handling personal data

1. Be open with people about what personal data they are collecting
2. Be open with people about how they use personal data
3. Be open with people about the way personal data is shared
4. Be open with people about the way personal data is secured
5. Explain to people how we make decisions about them using their data
6. Be open about their accountability mechanisms for misuse of personal data
7. Help people understand and influence how their data is collected and used
8. If collecting or using personal data, make their analyses and outputs as open as possible

How to use this report

The goal of this paper is to provide recommendations [for the REPLICATE project as a whole](#) and [for individual partners](#), to help guide the design of REPLICATE components. We believe that if partners implement these recommendations, they will be able to maximise the impact of data used and created by the project, while respecting privacy.

While the recommendations are specific to the REPLICATE project, many of the recommendations will be applicable to other smart city projects.

Summary of recommendations

1. Recommendations for the REPLICATE project	7
1.1 Commit to maximise the amount of data that is published openly, while respecting privacy	7
1.2 Identify one or more partners to coordinate the project's approach to data	7
1.3 Create and maintain an inventory of all the data assets being produced by the project partners and publish the inventory openly	8
1.4 Develop common data policy and licensing approaches to promote consistent data sharing and open data publishing	8
1.5 Develop common approaches to promote data integration and usability	10
2. Recommendations for individual REPLICATE partners	13
2a Respect privacy	13
2a.1 Carry out privacy impact assessments and publish them	13
2a.2 Create and publish a data usage policy	14
2a.3 Explain the benefits of open and shared data to participants	14
2a.4 Ensure affirmative consent is received and recorded, and a process in place for participants to withdraw this consent	14
2a.5 Design processes to give users control over data about them	15
2b Design for open	15
2b.1 Sharing ideas, issues, designs and thinking not just finished outputs	16
2b.2 Where possible, building on existing open source tools and open standards, and contributing back to these initiatives.	16
2b.3 Identifying datasets that can be released, and publishing them under an open licence	16
2b.4 Engaging data reusers and encouraging open innovation	17
2b.5 Explore potential business models which revolve around open data	17

1. Recommendations for the REPLICATE project

When talking to partners it became clear that the REPLICATE project as a whole needs to coordinate to ensure that a common approach is being taken to data. Ensuring consistency across the project will maximise the potential impact, scalability and success of the smart city platform. By taking these steps now, the project will be able to create solutions which are replicable and reusable beyond the project.

The following recommendations as listed in order of implementation with a set of suggested actions. Some actions could be carried out at the same event(s).

1.1 Commit to maximise the amount of data that is published openly, while respecting privacy.

The project should publish a declaration that it is committed to publishing open data by default, why this is the case and what benefits it expects it. Ensuring that all partners are aware of the value of data, and particularly the benefits of sharing and opening data in the context of cities will be key to implementing all the recommendations within this paper.

Suggested actions: Draft a short statement to be hosted on the project website and can be included in other project outputs. Partners should be familiar with the nature of open data and the value it has, possibly by engaging in ODI elearning.

Resources: [ODI's value of open data](#) and the [economic impact of open data](#) blog. [Online learning resources](#) (especially [unlocking value from open data](#)).

1.2 Identify one or more partners to coordinate the project's approach to data.

This could take the form of a single partner or working group for the whole project, or a lead within each of the pilot cities coordinating with a partner (or partners) from the ICT sub group. By taking responsibility for coordinating the data aspects of the project, they will be responsible for:

- coordinating partners to **complete and implement recommendations 3, 4 and 5.**
- ensuring partners are aware of, contributing to and using, the **resources created in recommendations 3, 4 and 5.**
- ensuring partners are aware of, contributing to and using, the **“Recommendations for REPLICATE partners building individual services”** below.

Suggested actions: Host a workshop or call involving partners across the project to agree on a suitable approach to data leadership. The resulting ‘data lead(s)’ should reach out to individual project partners about the importance and benefits of data management, data sharing and open data publishing.

1.3 Create and maintain an inventory of all the data assets being produced by the project partners and publish the inventory openly.

This inventory will likely build on an existing data management plan, it should be managed by the data lead(s) and be based on contributions from all individual partners. It should contain:

- A description of the data
- Who owns (or is the 'data controller' of) each data asset
- Whether it contains personal data
- Whether the data is commercially sensitive
- Whether it's a one-off data collection or an ongoing feed
- What the existing terms are under which it can be reused/shared?
- Identify whether data can and should be open/shared/closed, and explain the decision

To help make the decision on whether data should be open/shared/closed, the data lead(s) and partners should create and agree a common method for assessment. The full inventory list, including the decision made, should be published openly.

Suggested actions: Collect any existing data management plans and create a new document, preferably hosted online so that all partners can access it. Partners should contribute to it and it should be periodically reviewed. The data lead(s) should create a common method of assessment similar to the Defra Open Data Risk Assessment Tool, and seek feedback from partners. They could then host a workshop to help partners implement this assessment method for their own data.

Resources: data management plan tools: [DMPTool](#) and [DMPonline](#), [data spectrum blog](#), [ODI Maturity model - Part 1 data management](#), [Australian Councils toolkit](#), [USAID data development library process](#), Defra Open Data Risk Assessment Tool.

Depends on: Recommendation 1.2.

1.4 Develop common data policy and licensing approaches to promote consistent data sharing and open data publishing.

This could take the form of common documents that can be adopted and easy to use templates, and should include agreed approaches to the items listed below.

Depends on: All the following recommendations depend on recommendation 1.2 being met.

1.4a Create replicable terms around data licensing for procurement and contracting arrangements.

By creating a set of terms which can be included in all future procurement and contracting on the project, partners can ensure that any new components meet the requirements devised by the project. These terms should include the need to openly publish data, where appropriate, in machine-readable formats. They could also stipulate adherence to the common approaches stated in this section. Where possible, existing contracts and

deliverables should be reviewed and clarified in terms of these commitments to ensure those outputs are also as open as possible.

Suggested actions: The project ‘data lead(s)’ should draft a set of clauses which can be included in future contracts and seek legal advice to verify. This should be distributed to partners for comment to ensure it meets their existing procurement rules.

Resources: [procurement guide](#).

1.4b Develop a common data policy for describing how data is collected, used and shared.

Producing and agreeing upon a common high-level data policy relating to data produced and used in the project which would allow participants, data reusers and partners to understand how data is being used by the project in a consistent manner. This policy could then be linked from individual partners’ own more detailed data policies, ensuring consistency between the two.

Suggested actions: The project ‘data lead(s)’ should create a common policy, taking into account partners’ existing policies and test it with partners through a workshop. Partners would then endorse the policy and link to it from their own data policies. Partners could take step to review each other’s policies to ensure they are aligned with the high level common policy.

Resources: [ODI guide to writing a good data policy](#).

1.4c Develop standardised data sharing/licence agreements for integrating into the platform.

By providing a standard licence agreement for data shared with the REPLICATE ‘smart city platform’, the project can build a consistent legal infrastructure which will encourage reuse and innovation both within and beyond the project’s lifetime. This would be based on existing data agreements from partners and should conform to best practice guidelines around sharing of personal data.

Suggested actions: The project ‘data lead(s)’ should draft a data sharing agreement and seek legal advice. This should be distributed to partners for comment to ensure it meets their existing data sharing rules.

Resources: Concerning personal data: [ICO data sharing code of practice](#), [ODI draft checklist](#).

1.4d Choosing a standard open licence which data will be published under.

Choosing a single open licence to publish open data under will allow re-users to combine all the data produced by the project without having to check licensing terms. The licence used should preferably be an existing licence in common usage (for example, Creative Commons) to make the data compatible with other open data. Where institutions have existing licence requirements (UK government requiring OGL), partners will need to ensure that the standard licence chosen is compatible with this.

Suggested actions: Host a call or workshop to discuss licence requirements from individual partners and select the most appropriate licence.

Resources: [ODI publisher's guide to open data licensing](#), [Open data licensing. understanding your rights to use data](#), [Creative Commons' School of Open](#), [CLIPol compatibility tool](#).

1.4e Develop standardised communications and user engagement process that partners can follow when publishing data or launching new services.

By creating a standardised process to announce open data releases, data being made available on the platform or new uses or features, partners can ensure that the community and other partners are aware of the latest developments in the project and encourage them to engage. This could take the form of a checklist, which all partners are able to follow, and could involve announcements on social media, email newsletters and organising events. User engagement could be coordinated by a central partner within each city to explicitly promote the reuse of data. This framework and coordination should build upon any existing results dissemination plans and will ensure that any project requirement of this type is sufficiently met.

Suggested actions: The project 'data lead(s)' should draft a communications checklist which partners can follow when publishing data. Partners should decide whether to centralise coordination of user engagement within each city and select such a partner if desired.

Resources: [ODI guide to engaging with reusers](#), [ODI Maturity Model - part 3](#).

1.5 Develop common approaches to promote data integration and usability.

These should be informed by the ODI and Thomson Reuters' whitepaper '[Shareable by default: creating resilient data ecosystems](#)'. Specifically, partners should identify common approaches to the issues listed below.

Depends on: All the following recommendations depend on recommendation 1.2 being met.

1.5a Identifying how and where the data shared through the smart city platform will be stored

REPLICATE partners will likely generate and store data within the individual services they are creating. The project needs to decide whether the data that will be shared by these services with the smart city platform will be required to store data on the smart city platform itself or make it available in a more decentralised manner using, for example, open APIs.

Suggested actions: The ICT group within the project should evaluate the case for data to be hosted centrally in the platform or linked in a distributed manner. This should be circulated with city partners and a decision should be taken so that partners can plan accordingly. Design of the the smart city platform should take this into account.

1.5b Engage with existing open standards and capture requirements while developing a common data model and set of best practices for how to make the data produced easily interoperable

We recommend reviewing the potential data flows in the platform and identifying similar projects and the technical standards in use, including standard vocabularies, data formats and APIs. By using applicable open standards in the development of the common data model, the project can build on existing work being done in the field, avoid duplication of effort and ensure that the outputs of the project are interoperable. Where no standards exist, steps should be taken to ensure the data model is informed by engagement with potential re-users.

Suggested actions: The project 'data lead(s)' and ICT group should work with partners to explore existing standards which could be used, potentially during workshop. Design of the the smart city platform should take this into account.

Resources: [W3C standards](#), [UK government open standards principles](#), [BSI smart city standards](#), [Shareable by default: creating resilient data ecosystems](#).

1.5c Identifying potential common approaches to anonymisation and aggregation, and decide on which approach(es) will be used for data that will be shared or openly published

By identifying common approaches to anonymisation and aggregation, partners will be able to ensure that best practices are being followed to preserve the privacy of participants and that the data produced is interoperable. Through using the same techniques, partners will be able to coordinate communication with participants and make it easier for data reusers to understand how the data has been processed. In addition, by considering common methods partners will be able to more accurately judge the risk of de-anonymisation where participants are described in multiple data sets and will be able to actively counteract this risk.

Suggested actions: Host a workshop for partners working with personal data to explore the anonymisation requirements of the project. The output should be shared with other partners and published openly.

Resources: [UK Anonymisation Network resources](#)

Depends on: Recommendation 1.3.

1.5d Identifying common methods for making open data available and reusable

This could be through publishing on an existing open data platform, such as the [Bristol open data platform](#), or through a centralised catalogue listing of data published in a more distributed way, on partner websites for instance, or through the smart city platform itself using either of these methods. Whichever approach is taken, it should be applied consistently across the project and will depend on recommendation 5a. To ensure that data is reusable, the published data should be assessed and certified. This could be done by specifying a minimum standard for ODI certificates for instance, to ensure all data publishing is meeting best practices.

Suggested actions: The project ‘data lead(s)’ should create a case for publishing open data on a specific platform and share with partners. Design of the the smart city platform should take this into account.

Resources: [ODI Open data certificates](#)

Depends on: Recommendation 1.3.

2. Recommendations for individual REPLICATE partners

Each of the REPLICATE partners building individual components will have to work closely together to define the common practices that surround their use and sharing of data, as explored above. However, they will be generating different types of data from different new or existing systems, and will therefore face some different considerations when building their components.

Each partner is responsible for ensuring that the privacy of their users is respected. In addition, partners should be ensuring their components have maximum impact by working in the open – publishing open data, open source code, learnings and engaging communities.

2a Respect privacy

[ODI data infrastructure principle 3:](#)

“In the most impactful and valuable data infrastructure, openness is maximised but what is private remains private. Different countries have their own data protection legislation and social contracts, which need to be adhered to. To build trust organisations using personal data should also be open with people about how they use and share that data.”

From our discussions, it is clear that many, if not all, REPLICATE partners collecting and using personal data have already thought through the issues surrounding such activities. They will all be compliant with existing laws, often with existing procedures in place. However, we recommend that partners go one step further in ensuring that they are using best practices which build trust through openness around the collection and use of personal data. It is also important to ensure that partners are able to maximise the amount of open data published, for example not just keeping data closed because of potential privacy issues without considering potential mitigation. This will also prove useful in ensuring the sustainability of the platform once the General Data Protection Regulation comes into force.

To do this partners should use the [ODI's openness principles for organisations handling personal data](#) and complete the accompanying [draft checklist](#) to ensure each project is following best practice. While doing this, partners should particularly pay attention to the following steps.

2a.1 Carry out privacy impact assessments and publish them.

Partners should carry out and publish privacy impact assessments to help them and others understand the implications of their service for the privacy of individuals. They should use the results to guide further development and act as a reference point for customer queries. An effective privacy impact assessment should be dynamic, in that it should be updated to reflect significant changes both from changes to the project itself and, for instance, exterior technological developments or information releases increasing re-identification risks. These will become mandatory under GDPR which is scheduled to come into force in May 2018.

Suggested actions: Each partner should carry out privacy impact assessments on their services, based on the data they intend to produce. These should be periodically reviewed.

Resources: [ICO privacy impact assessment code of practice](#), [ICO privacy impact assessment code of practice additional guidance](#) (download).

Depends on: Recommendation 1.3.

2a.2 Create and publish a data usage policy

Partners should create and publish a data usage policy, which should link to the common data policy created in recommendation 4b. It is particularly important that this data policy references the collection, management and sharing of personal data. It should also be written in clear and accessible language so that any user can understand it and how data about them will be collected, used and shared. Partners should take steps to make this more accessible through use of different media and by user testing.

Suggested actions: Each partner should publish a data usage policy based on the ODI guide to writing a good data policy and linking to the common data policy.

Resources: [ODI guide to writing a good data policy](#), [ODI draft checklist](#).

Depends on: Recommendation 1.4b.

2a.3 Explain the benefits of open and shared data to participants.

This should help participants make an informed decision to share their data - especially as the benefits of sharing through REPLICATE will likely be felt by them as well as commercial partners. Efforts should be made to make these as accessible as possible to a range of participants.

Suggested actions: Partners should develop a short, clear articulation of the benefits of sharing data as it relates to the REPLICATE project. This could be done in coordination with the 'data lead(s)' and other partners during one of the workshops proposed.

Depends on: Recommendation 1.1.

2a.4 Ensure affirmative consent is received and recorded, and a process in place for participants to withdraw this consent.

Partners should take steps to ensure they are following best practices when seeking consent from participants. By taking steps to ensure that participants are fully informed and aware of their legal rights, partners can build trust in their solutions and the REPLICATE project amongst participants and the wider community.

Suggested actions: Partners should ensure they design systems with affirmative, verifiable consent, potentially in coordination with the 'data lead(s)' and other partners.

Resources: [Projects by IF Design Patterns for Consent](#), [Projects by IF designing for new digital rights](#), [ICO privacy by design](#), [ICO GDPR considerations](#).

2a.5 Design processes to give users control over data about them

Partners should design processes that will give users wide control over data about them, in line with upcoming GDPR requirements, including around correcting data, removing data, accessing their own data and, for portability purposes, being able to share data about them with named parties. Partners could do this through open APIs that allows users to share data to open standards including both human and machine-readable formats. This method would create the opportunity for open innovation as well as complying with EU regulation.

Suggested actions: Partners should ensure they design systems with correction and access to personal data in mind. A workshop could be held by the data lead(s) to discuss what access might look like consistently across the project, and to share learnings from the design of such systems.

Resources: [Projects by IF designing for new digital rights](#), [ICO GDPR overview](#), [Open Banking standard](#).

2b Design for open

[ODI data infrastructure principle 1:](#)

“Open data, open culture, open standards, open source and collaborative models build trust, reduce cost and create more value than other approaches. Being open improves quality as more people can contribute to the outcome and increases the number of connections that can be made. Data benefits from network effects: it creates more value as more people use, contribute to and maintain it.”

From our discussions, it is clear that REPLICATE partners are keen to maximise the impact of the project, their components and the data that is generated. While much of this will come from effectively working together to develop common practices, some of it will originate from people and organisations outside the project.

By working more openly – building on the work of others and empowering them to build on the work of REPLICATE – partners will be able to ensure the project has sustained impact. Working in the open also makes it easier for partners to collaborate and keep each other up-to-date on progress.

To help partners ensure this impact, we recommend partners work in the open as much as possible by:

2b.1 Sharing ideas, issues, designs and thinking not just finished outputs

For example, this can be done by starting a project blog, sharing notes from internal and external workshops, publishing project presentations, opening up access to webinars and a range of other options.

Suggested actions: Partners should begin to publish blogs and open up non-confidential notes to track the progress of the work being done on the project.

Resources: [Mozilla 'Working open' guide and resources](#), [ODI blog on Open Culture](#), [Defra presentation and blog](#).

2b.2 Where possible, building on existing open source tools and open standards, and contributing back to these initiatives.

By building on open source tools and standards, partners are able to make best use of existing work in the field. Where these resources are available and appropriate, partners can engage an existing community who will be able to provide support. By contributing back to projects, either through code or feedback, the project will be able to maximise its ongoing wider impact. Partners should take steps to ensure that components which are not continued or supported after the project are made available as open source projects.

Suggested actions: Partners should work to include existing open source software and open standards into their development process.

Resources: [Mozilla 'Working open' guide and resources](#), [GitHub open source guides](#), [UK government guide to open standards](#).

Depends on: Recommendation 1.5b.

2b.3 Identifying datasets that can be released, and publishing them under an open licence.

All REPLICATE partners should be able to release some open data, either non-personal operational data or personal data which has been suitably anonymised, aggregated or where permission has been explicitly granted.

Suggested actions: This process leads on from many of the previous recommendations.

Resources: [ODI publisher's guide to open data licensing](#), [Open data licensing, understanding your rights to use data](#), [Creative Commons' School of Open](#), [UK Anonymisation Network resources](#).

Depends on: Most prominently recommendations: 1.1, 1.3, 1.4d, 1.5c, 1.5d.

2b.4 Engaging data reusers and encouraging open innovation.

Partners should aim to build a community of data users and reusers who are aware of the data, use it and provide feedback. This can be done through focused events (such as hackdays, meetups, challenge series) and open consultations.

Suggested actions: Partners should work to identify challenges in their cities that might be solved using the data created by the project. Efforts should be made to engage local communities through holding external events.

Resources: [Open Data Challenge Series handbook](#), [open Innovation e-learning module](#), [ODI guide to engaging with reusers](#), [ODI Maturity Model – part 3](#).

Depends on: Recommendations: 1.4e, 2b.3.

2b.5 Explore potential business models which revolve around open data.

Partners producing new services and updating existing services should consider what role open data could play in their business model. Partners should consider what operational models already exist for services that publish open data, including those for which publishing open data is a core part of the business. They should also consider the benefits publishing open data might have beyond direct revenue, looking at wider economic, societal and environmental impact, for example.

Suggested actions: Carry out research into existing open data businesses and models. Host external workshop on findings and invite collaboration with research partners.

Resources: [ODI guide on how to make a business case for open data](#), [ODI case studies](#), [Open data means business](#), [Open enterprise](#).

Acknowledgements

Produced by the Open Data Institute at the request of Bristol City Council for the REPLICATE project.

This activity has been developed under the REPLICATE project which is funded by the European Union's HORIZON 2020 Research and Innovation programme under grant agreement No. 691735.

Authors

Jamie Fawcett
Leigh Dodds
Tom Sasse
Peter Wells

Delivered

16 June 2017

Licence

Distributed by the Open Data Institute and Bristol City Council under a [Creative Commons Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/).

