Knowledge & opinion > Reports >

Professional surfers carrying their surfboards while going to the sea, professional surfers in black diving suits ready to surf walk to the ocean, close up of surfboard with surfer on background

Prototyping with open sports data (report)

Mon Apr 9, 2018

Share

Built environment and housing **Health and physical activity** **Cities**

Three prototypes exploring the potential of local, open data to support communities and enhance the experience of peer-to-peer accommodation platforms

Olivier Thereaux
Former ODI staff member: Head of Research and Development

For the past few months, as one of our innovation programme projects, we have been exploring the role of data in understanding and measuring the impact of – and informing the debate about – the emerging peer-to-peer accommodation sector.

Most of the exploration into the data itself focused on data about the sector and monitoring its various impacts on local areas, but we also wanted to look at it from the other side of the equation: could data, especially open data – covering other sectors and areas of society – benefit both communities and actors within the peer-to-peer accommodation sector?

By asking this question, we are working to understand how data can support a better experience for both peer-to-peer accommodation sector users and for the communities they visit. We think that by working with data about local communities and local businesses, peer-to-peer accommodation platforms can support and engage with local communities and create mutual economic opportunities. We have looked at the specific example of peer-to-peer accommodation platforms working with data about physical activities, however, there are other data types that could be connected. These range from local public services and community activities to shops and cinemas. There are also different platforms and services which could create more opportunities and a richer environment for innovation, social and economic value.

Early in the project, our team thought we should explore how existing open data – a lot of which is inherently about places – could be used by platforms to enrich the experience of peer-to-peer accommodation. One of our motivations was to address the fear we were hearing from users and local authorities that communities are being impacted negatively by peer-to-peer accommodation platforms, or the people brought into their neighbourhoods through these platforms. These fears included noise pollution, dangerous or careless behavior, disruption of local communities and displacement of inhabitants, among others. We wondered if this effect could be minimized by making platform users more aware of their surroundings, which in turn would help them bolster the local economy, participate more in local life and help create a more positive environment.

Stewardship v integration

Unlike data gathered and maintained directly by peer-to-peer platforms about the neighbourhoods in which they offer accommodation, we started to realise that integrating with third party data has its challenges as well as advantages, and requires new approaches. Integrating open data maintained by third parties into peer-to-peer platforms would mean that different platforms could access, use, and even contribute to any given locally focused dataset in completely different ways. And so we set out to develop several prototypes around a single dataset to explore this hypothesis.

It is worth highlighting that these same benefits and challenges might apply to any service integrating with any open data, and so these prototypes will prove useful to surface those issues for other readers, regardless of sector.

Augmenting the experience with physical activity data

Based on user interviews, we considered a number of sources of data that could be considered useful in tailoring trips to individual needs, including data about bus stops and times, local restaurants, independent shops, local event venues and things to do, as well as when an area is busy versus when it is not. There might also be an interest in data about the area, for example census or economic data, or crime statistics. Some of this data already exists as open data in certain cities and councils, but it probably hasn't been collected in a standardised, usable way across the different regions of the country.

Instead, we focused on physical activity data, partly because we had seen anecdotal interest for the mixing of the two, as illustrated in this Twitter exchange:

Physical activity data is inherently local and becoming broadly available across geographies. Unlike practical or transactional data – such as waste collection or the location of bus stop – data about sports and activities lends itself well to being used creatively by peer-to-peer accommodation platforms in ways that could open up opportunities for innovative added value to their services. This gives us a richer variety of uses for our prototypes than if we had worked with other open data sources such as bin collection or parking data.

Finding and choosing the most useful dataset

At the ODI we have been leading a sector programme on physical activity data called OpenActive, through which a lot of open data on activity opportunities

(classes, outings, etc) are being published in a decentralised way by organisations and communities, using a standard we helped develop. For this set of prototypes, however, we were more interested in the location of facilities for sport and physical activities than in the activity opportunities themselves. For the purpose of our prototypes, the Active Places Power open dataset published by Sport England seemed most appropriate: with its focus on facilities, the Active Places Power set includes more consistent data about geographical location of the facilities than in the OpenActive datasets, where precise latitude and longitude information were made optional to make publishing more straightforward.

The Active Places Power dataset is focused on facilities around the UK and is managed and produced by Sport England as part of their work to assess those facilities, and is a readily available dataset for exploring the questions we set out in this report. However, there is a much wider set of locations where physical activities take place, such as parks and public spaces. Over time, we hope that OpenActive datasets will help reveal a broader set of locations outside sports facilities, and grow to include identifiers for those places.

Three prototypes

We ended up developing three distinct prototypes, each interacting and integrating with the same dataset:

1. How might a peer-to-peer accommodation platform integrate with an open dataset to offer a richer experience, while controlling risks of using data they do not directly gather or maintain?
2. How might a peer-to-peer accommodation platform use this same dataset with an approach that is more locally focussed, more actively engaged with the local community and responsive to local feedback?
3. How might a more specialised peer-to-peer accommodation platform curate the same dataset, to add value beyond the display of information, but also as a vector for business development and partnership building – for example, to broker services between sports facilities and its accommodation seekers?

Prototype one: Integrating with open, local data

Prototype two: Integrated with the local data ecosystem

Prototype three: Finding value in and adding value to data

Conclusion

In this report we explored, through practical prototyping, how peer-to-peer accommodation platforms could use, integrate with and curate available open data sources to enhance the experience they offer their users. While we focussed on the

peer-to-peer accommodation sector in this instance, we would like to point out that the same challenges and benefits would apply to any service wanting to integrate with open data, regardless of sector.

View the conclusion

Related

Recent related activities

BLOG, CASE STUDY 24/08/2022

Helping the physical activity sector use open data to get more people active

We worked with Sport England to develop OpenActive – a community-led data access initiative to get more people active using open data

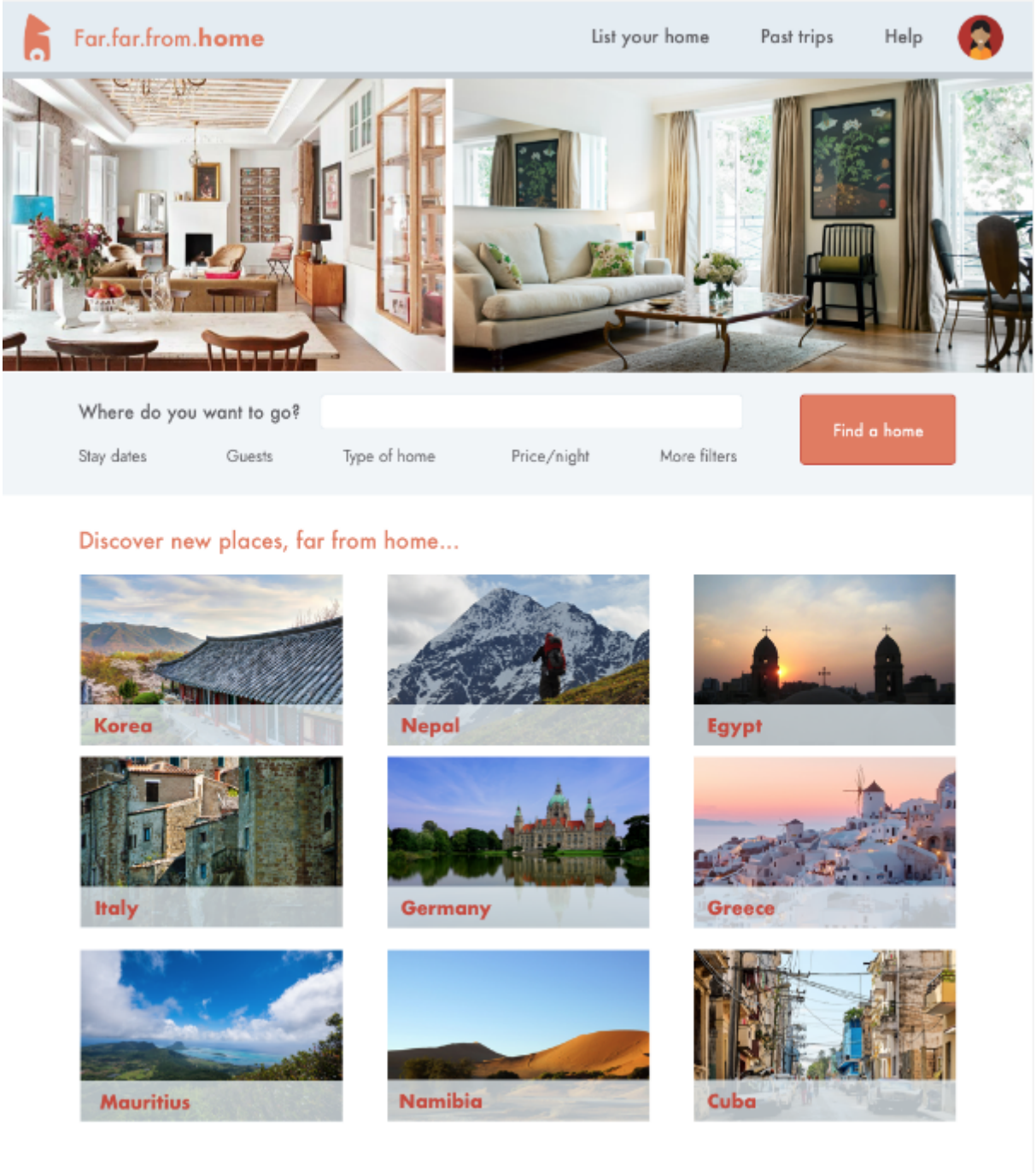
DATA PUBLISHING AND USE, HEALTH AND PHYSICAL ACTIVITY, INNOVATION

Prototype one: a global platform integrating with open, local data

The platform “Far.far.from.home.com” adds value to their original offer by being more than a simple peer-to-peer property broker. They are a platform that would like their users to get the best experience possible out of the available rentals, and through their own research they have discovered that users want to take part in local community life and activities. In order to do that, they’ve decided to highlight local sports and activity centres.

The popular, worldwide platform “Far.far.from.home.com” is constantly striving to add value to their original offer by being more than a simple peer-to-peer property broker. They are a platform that would like their users to get the best experience possible out of the available rentals, and through their own research they have discovered that users want to take part in local community life and activities.

In order to do that, they’ve decided to highlight local sports and activity centres.



This data is made available by Sport England as a publicly available dataset. In this case the data is made available as a complete data file. Without a searchable API provided by Sport England anyone who wants to search the data needs to download the datafile and store it in a local database. For the latest corrections and updates to the data, the platform *can* call an API; the corrections and updates are merged with the platform's local database.

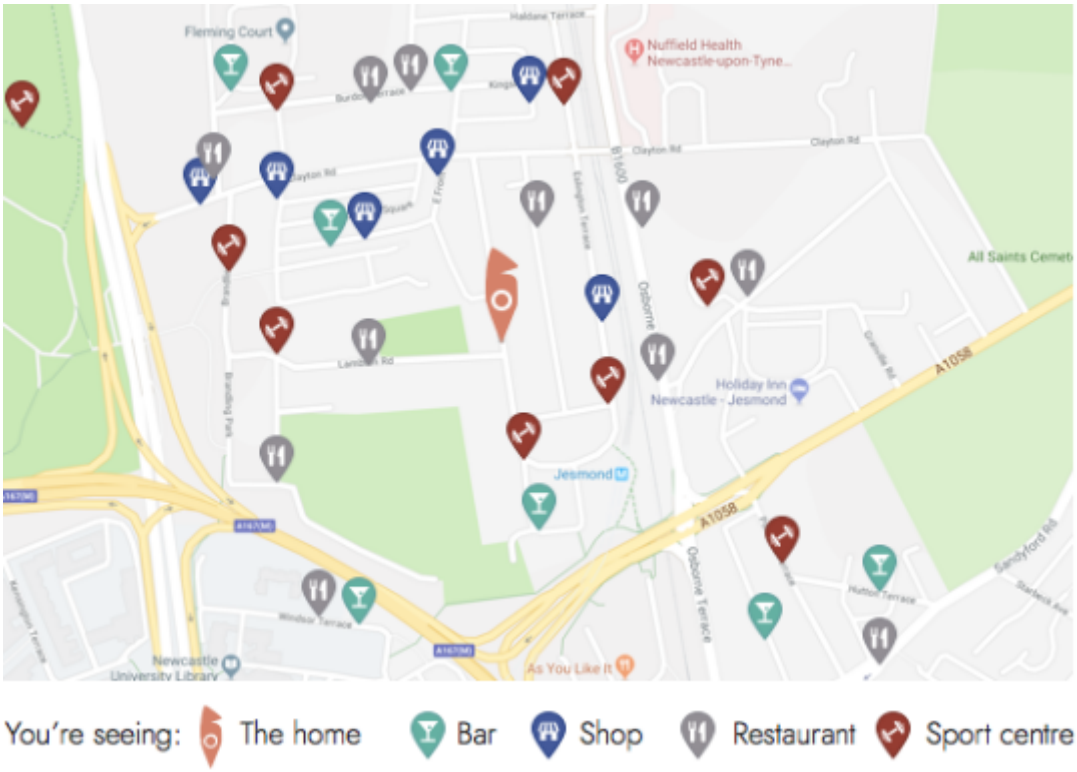
The data for each site offering physical activities contains the following and more:

- Location information such as address and postcode for searching, and latitude/longitude for placing on a map
- Access information such as provision of disabled parking, toilets and changing facilities

- A list and details of facilities such as gyms, pools, football pitches, tennis courts and many more facility types

The user experience

Each property has a details page that includes a ‘Nearby’ section. The property would be highlighted on a map, along with shops, bars, restaurants, and sports centres.



The map would be interactive, with users able to click on each ‘pin’ to open a panel of detailed information about the facilities, taken directly from the dataset.

You're seeing: The home Bar Shop Restaurant Sport centre

The best Gym 123 Eslington Terrace, QW32 9ER Newcastle

Opening times:

- Monday to Friday: 7am - 10pm
- Saturday: 6am - 11pm
- Sunday: 6am - 11pm

Open every day including bank holidays

Fees:
£12 - £90 depending on activities

Amenities

- Fitness classes
- Treadmills & static bikes
- Yoga classes
- Weights room
- Pool
- Water rowers
- Climbing wall

Responsibility for data quality and impact on user experience

At this point the platform might want to consider who – from the user's point of view – is considered liable or responsible for the accuracy of the data.

In the case of a large platform with millions of users worldwide, the possibility of adopting and using open data can feel like it carries a significant risk. Incorrect data shown as information on the platform could lead to some unpleasant situations: what if the address for a facility has changed, and the platform recommends its users go jogging in what is now a toxic waste plant? As they consider the potential value of this dataset and how it might benefit users, the platform's lawyers and product managers will undoubtedly consider these potential outcomes, affecting how they might decide to mitigate risks.

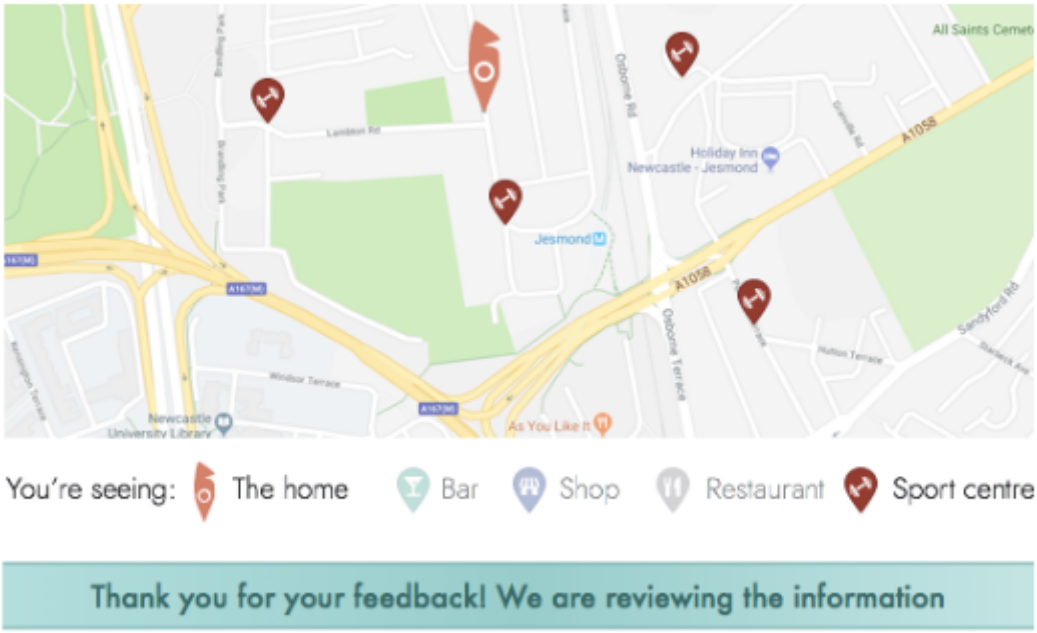
The first mitigation strategy is for the platform to clearly state the provenance of the data. Doing this right can be a challenge: in the typically information-rich interfaces of online platforms, it can be hard to get people to notice, let alone understand, that some information comes from, and is under the control of, a third party. But not only is stating provenance a fair way for the platform to give credit to the source of the open data, it can, when done right, also help direct comments to the right channels should the users of the platform notice missing or incorrect data. This is why, for example, the BBC visibly credits the source of information when integrating open content and open data in their online services.

Demonstrating provenance does not fully address the concern that the platform may be showing incorrect or otherwise problematic information based on their integration of this open data.

Here the platform has a choice of mechanisms to adopt. Assuming that the platform adds an interface for its users to report incorrect information:

- They could simply hide incorrect data to minimise the impact on others, and the effect on their reputation for quality
- They could inform the user that the data is managed by someone else and redirect them to the stewards of the data, in anticipation of receiving a correction via the Sport England API at a later point. In effect this conveys that they are exempting themselves from liability and responsibility
- They could correct their local copy of the data, or update it from the source, where it might have already been fixed
- They could take it upon themselves to make sure the data is fixed at the source, assuming the source has feedback mechanisms to support this. We explore this in prototype two
- They can also provide a mechanism for integrating corrections supplied to them back into the original data source, thereby connecting the user and the third party

Adding contextual information can help to build trust and create transparency. This includes indicating when the data was last updated, and using high quality sources, such as those published at source by organisations like Sport England, or by checking the provenance of data supplied by intermediaries.



In this example, Far.far.from.home has opted for a mix of the above: the user can 'Report information as incorrect', and the platform automatically hides that particular location while they make efforts to rectify the mistake, potentially at source. This way other users will not see the incorrect information and the platform won't get in trouble for displaying inaccurate data to users. Platforms might also have policies in place to avoid abuse between competitors, for example where a company repeatedly flags a competitor's location as incorrect in an attempt to gain more business for themselves.

What happens next, as illustrated here, is unknown to the user: they do not know if the feedback is sent to a member of platform staff to check, passed on to the third party to check, or permanently deleted on Far.far.from.home's database. Is the user expected to notify the third party themselves if they wish the data to be accurate for future users, or users of other platforms? In the second prototype (detailed below), we will explore a different approach.

Prototype one walkthrough: https://14dim6.axshare.com/#c2

Summary

Benefits of integrating third party data

- Third parties have specialist knowledge and understand sophisticated domains, beyond accommodation – the peer-to-peer platform can concentrate on what they do best while gaining value from the third party's expertise and services
- The user has access to information that they wouldn't see or know about in other circumstances, for relatively little effort on the part of the platform
- Integrating data enhances the visibility and utility of that data to end users, giving an incentive to data stewards to publish more and keep what they

publish up to date, which in turn makes it more useful and usable for both the platform and others

Challenges to integrating third-party data

- Platforms will need to synchronise with open data on a regular basis to keep up to date and provide the ability to search quickly
- The data is not managed by the platform, leaving the possibility of presenting errors and inconsistencies for which the platform may appear liable. Furthermore, the platforms can't directly change the source data in the case of mistakes and may encounter difficulties keeping locally corrected data both correct and synchronised with the original source
- Different third parties make their data available in different ways, using different formats and API conventions. The platform will need to take different technical approaches depending on the data provider
- Without providers adhering to standards, it would be difficult to combine data. For example, different providers may use different identifiers for locations, or use different data models to describe the same domains
- Platforms have no control over when and how the third party might change the format or availability of data that they rely on, but their customers might come to rely on and expect that data to be present. The platform will need to have contingencies in place for unexpected interruptions to the provision of data, or its gradual obsolescence if the third party stops investing in its maintenance

We have attempted to explore the challenges of feedback and corrections in our second prototype.

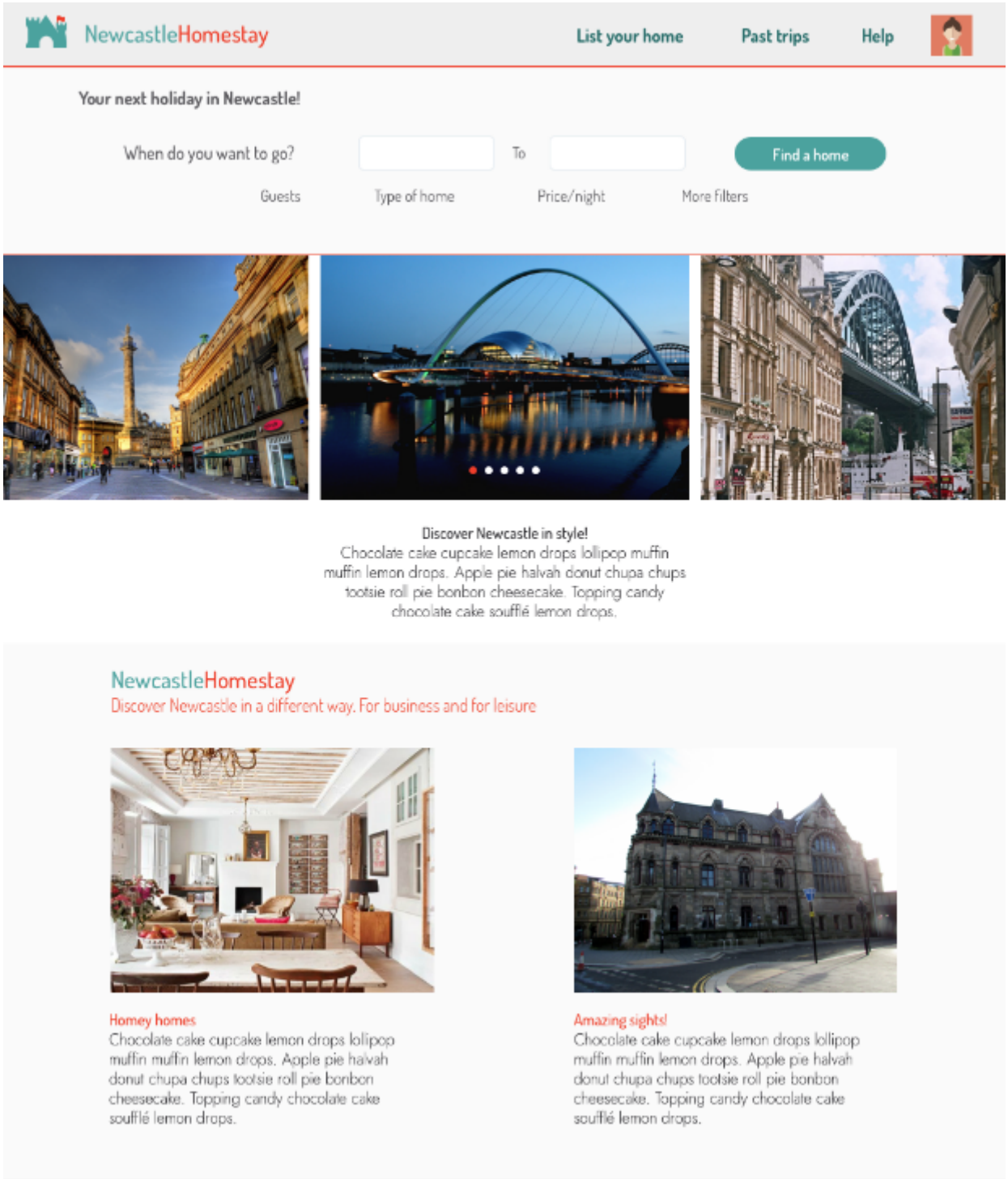
Go back to main page View prototype two

Open Data Institute, 5th Floor, Kings Place, 90 York Way, London N1 9AG
Privacy policy Contact us

Prototype two: a platform integrated with the local data ecosystem

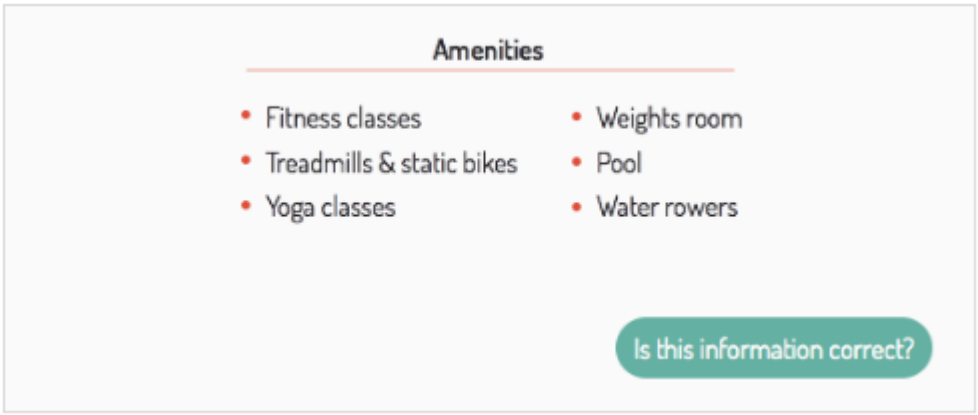
This platform enables users to become part of the local data ecosystem for everyone's benefit: the users can input their own local knowledge, the platforms benefit from more accurate information, and the open data becomes more accurate and useful for everyone.

What distinguishes this from the first prototype is that the platform considers itself more integrated in the local data ecosystem. We think this is more likely if either the platform has a more local focus and has built good relationships with the people who provide the data, or if there were good data communication standards and a technological process in place to manage amendments.

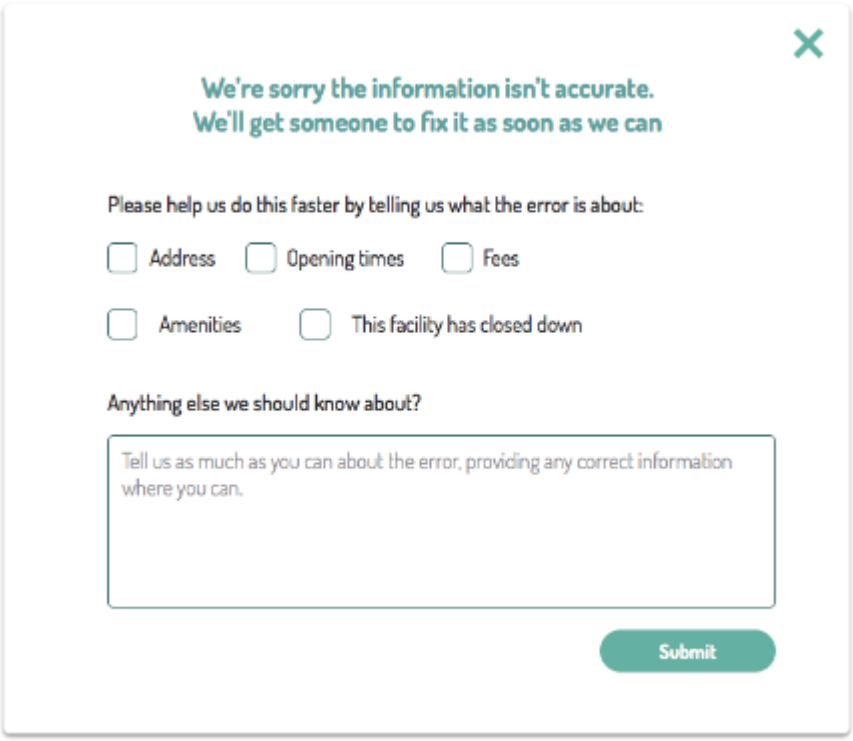


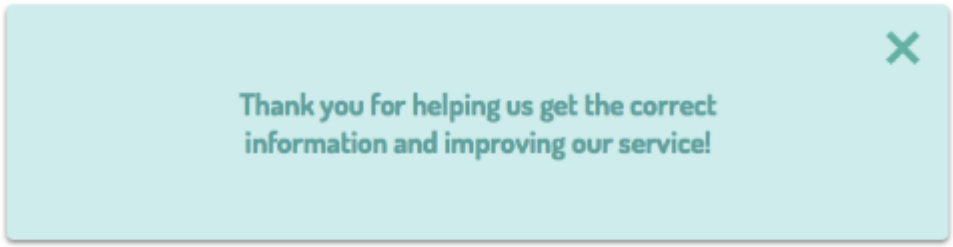
This way the platform and users become part of the local data ecosystem for everyone's benefit: the users can input their own local knowledge, the platforms benefit from more accurate information, and the open data becomes more accurate and useful for everyone.

The user experience is much the same, but in this case when the user finds an error, the platform provides an interface to enable the data to be corrected, thereby effectively crowdsourcing quality data.



This way the correction is captured at source, in the moment, and sent directly to the third party: the platform acts as a bridge between the user and data providers.





The platform can choose how to track the integration of the correction and communicate it to the user.

Prototype two walkthrough: **https://t0v9qj.axshare.com/#c=2**

Summary

Benefits of feeding back corrections

- Develops relationships between data providers and platforms, encouraging dialogue and creating an incentive for the data provider to keep providing quality data
- By feeding corrections back, the platform ensures that the data at the source remains usable in the long run. The alternatives are much less desirable. Simply hiding incorrect data misses the opportunity of it being corrected, and correcting only the data in the local copy held by the platform means that the two sets will diverge over time. This means that the platform cannot benefit from updates from the source, and has to maintain a parallel, but flawed, version of the dataset forever.
- Multi-way conversations lead to convergence on standards and integration, which benefits the data ecosystem as a whole and can encourage further innovation by other platforms
- Both the platform and users are involved in a transparent way, increasing understanding of data's role in delivering a service
- Shared responsibility for data quality reduces the costs involved for any single party in keeping data up to date and relevant

Challenges to feeding back corrections

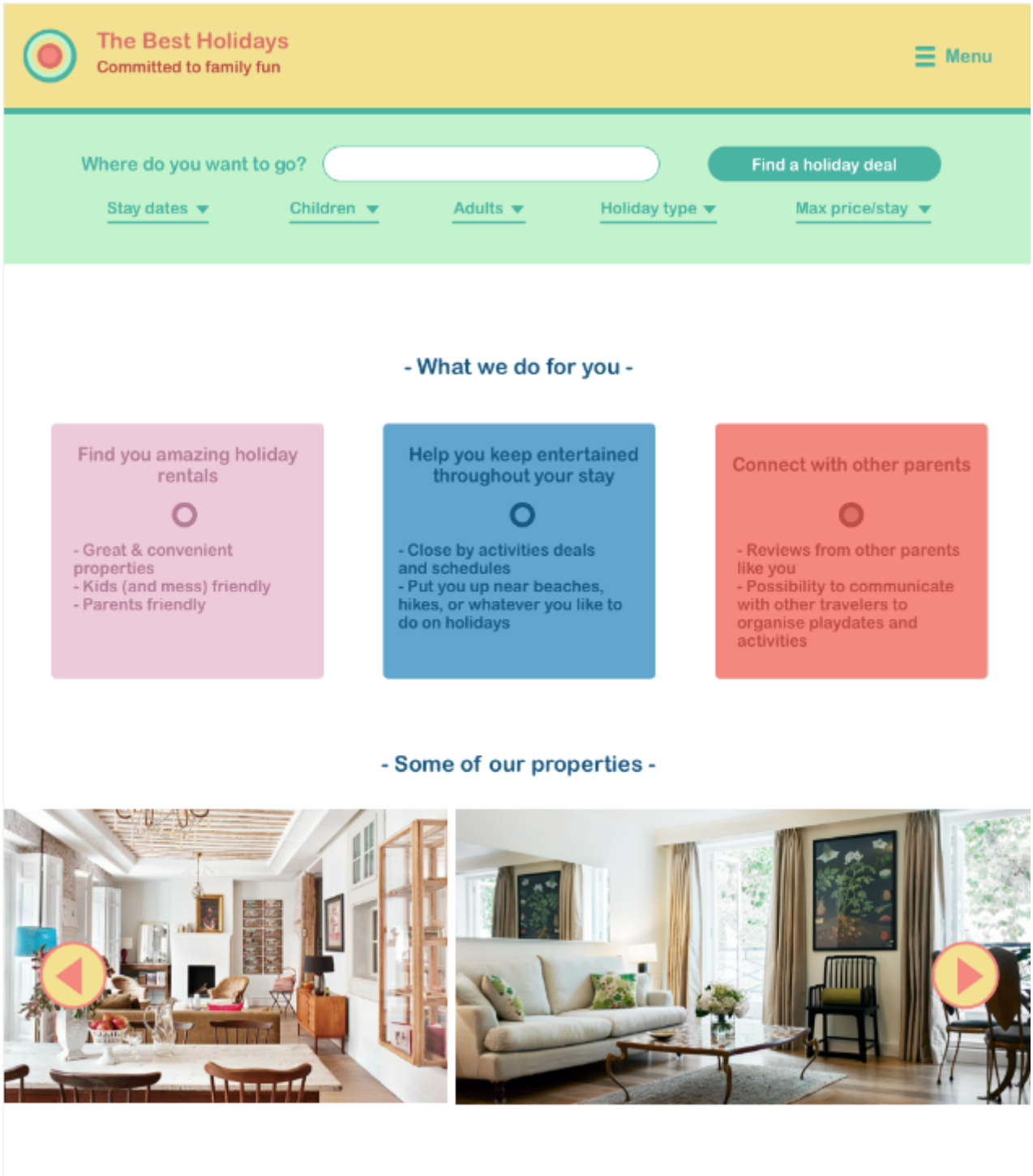
- Integrating with third parties necessitates building additional interfaces between the platform and third party data providers, and this takes additional resources
- The ability to feedback corrections requires data providers to supply feedback interfaces. This is likely to make additional work for them and may only be worthwhile if useful and accurate corrections are likely to be provided through the platforms
- Feedback may take time to be integrated into the third party data but users may expect changes they suggest to be displayed immediately. Platforms may need to amend their local copy of the data while waiting for the correction to work its way through into the original
- In some data ecosystems, there may be intermediaries between the platform and the original data source. To feedback corrections, feedback interfaces would have to be created at every step of the data value chain

Back to prototype one | Back to main page | Prototype three

Prototype three: finding value in and adding value to data

This prototype explores how a peer-to-peer accommodation platform called The Best Holidays adds further value on top of an open dataset by curating it and augmenting it for their own users.

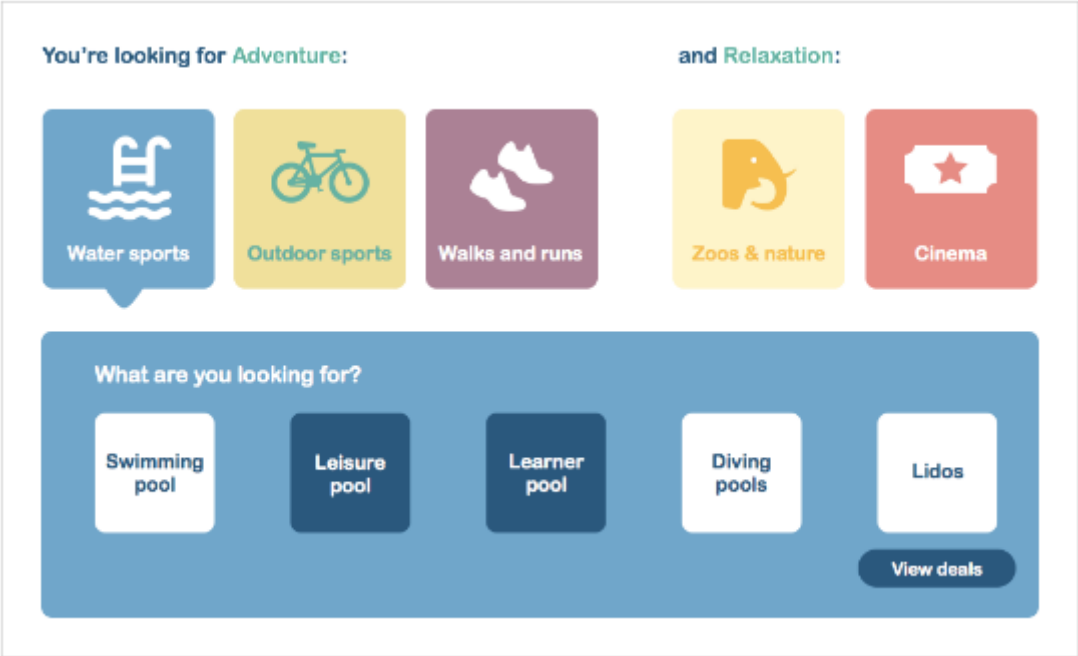
Our third prototype explores how a peer-to-peer accommodation platform called The Best Holidays adds further value on top of an open dataset by curating it and augmenting it for their own users.



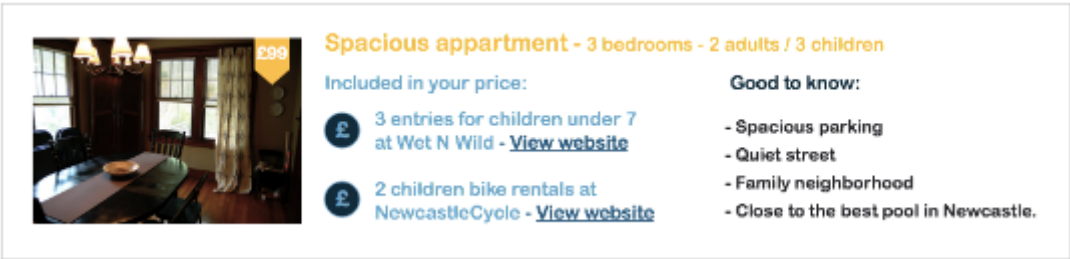
The Best Holidays caters to families. Their business focus is to highlight accommodation in quieter areas away from busy roads, or whose owners have considered issues such as storage for buggies, space for travel cots and safety features such as stair gates. Another important consideration for families are suitable activities, such as alternative plans for rainy days.

Looking again to the Sport England dataset, The Best Holidays can identify local sports facilities that would suit a family. For example, swimming pools that are defined as ‘Leisure pools’, which are described in the accompanying API documentation as “primarily designed for informal recreational swimming and may include flumes, slides, beach areas, water jets, and wave machines”. They can also identify ‘Learner/Teaching/Training’ pools, suitable for younger children.

The Best Holidays adds their own unique value to an open dataset by selecting only data that is relevant to their business model, and presenting it in a way that makes best sense to their own users.



In this example, besides showing the Sport England dataset information, they have shown information relevant to parents that would benefit the user, such as proximity to child-friendly restaurants, neighborhood quietness, and closeness to parks and museums.



This data is obtained by cross-referencing property listings with points of interest in OpenStreetMap (parks, restaurants, museums, soft play areas, etc) and integrating that with property listings.

Finding opportunities in data

By exploring data that is collected and managed by a third party who knows a domain well, it is possible to discover opportunities and connections that are unique to The Best Holidays without incurring the cost of data collection. In other words, the data facilitates innovation, business development and partnership building, and strengthens relationships in the communities in which The Best Holidays operates.

The Best Holidays approaches those sites identified as providing family-friendly facilities or activities and agrees special discount or package deals. This serves the double purpose of bringing business to the sports facilities and making the accommodation platform more attractive to families.

The prototype presents this concept from the user perspective. The Best Holidays has already used the dataset to identify partner facilities, and has agreed special package deals for the families using this family-friendly accommodation platform. These deals are offered to the user through the website.

Prototype three walkthrough: https://qm2t2r.axshare.com/#c=2

Summary

Benefits of showing curated information

- Gives platform a competitive advantage: they can provide services that are more enriched with relevant information than other platforms
- Allows platforms to create more targeted, distinct services
- Improves services for end users
- Reduces the cost and risks of adopting a dataset maintained by a third party – through curation, less of the data gets displayed, and the likelihood that the

Conclusion and recommendations

In this report we explored, through practical prototyping, how peer-to-peer accommodation platforms could use, integrate with and curate available open data sources to enhance the experience they offer their users.

In this report we explored, through practical prototyping, how peer-to-peer accommodation platforms could use, integrate with and curate available open data sources to enhance the experience they offer their users.

While we focussed on the peer-to-peer accommodation sector in this instance, we would like to point out that the same challenges and benefits would apply to any service wanting to integrate with open data, regardless of sector.

The prototypes illustrate how this integration can be achieved, at least technically, in a very short amount of time. It also helps us illustrate some of the challenges such an integration would create.

On the one hand, useful local data exists and much is available as open data for anyone to access, use and share.

But there are costs. Using open data requires an investment of developer time to deal with the incoming third party data, particularly when working to integrate a lot of different third party sources that have not been built according to the same standards and use different APIs.

Additional work is needed to integrate feedback loops between data providers and platform users. They may not need to collect or curate the data themselves, but integrating other datasets could divert platforms from their core functionality and require them to become knowledgeable in domains they might not be comfortable with or benefit from right away.

However, for some platforms, using data to invest in more targeted business models could give them the edge over bigger platforms. It could help them find and cater to niche users and create a loyal audience that can rely on their richer offer. Curation of information and delivering to targeted audiences could be a way to overcome competition in the long term and change the peer-to-peer landscape over time.

If done well, displaying and curating data may alleviate some fears we've heard concerning peer-to-peer platforms not taking into account local communities or

creating a bubble of tourism in certain areas. While our prototypes focused on sports and physical activity data, interacting with other local council data (regarding bins, crime statistics and noise statistics for example) and local area data (local restaurants, artisans, small shops, neighborhood activities, etc.) could help with those issues.

Drawing data from collaborative platforms like Open Street Maps could give local businesses the opportunity to ensure they have more control over the data they show, and that that information is up to date. Local councils and business information districts could offer support to do this, while other services provide mechanisms for community groups to provide data, such as the accessible locations data provided and crowdsourced through Wheelmap.

However, a greater variety of data sources brings challenges which place a heavy burden on platforms that want to make use of this data, such as:

- handling multiple data formats and API approaches
- integrating data that uses different identifiers
- dealing with differences in responses to corrections and updates

Our prototypes used only open data and therefore did not have to contend with the additional compliance challenges that would arise if different datasets were made available under different terms and conditions. The technical challenges are hard enough.

Intermediaries can thrive in such a complex environment, offering pre-cleansed, standardised, integrated and reliable data, and adding sufficient value to data that is provided for free at source that services further up the data value chain are prepared to pay for it.

From this prototyping exercise, we can highlight a number of recommendations for data publishers, user-facing services and intermediaries:

Data publishers:

- Make it easy for people who use your data to get in touch with you and talk about their plans – this will help you ensure the data is being provided in a useful way and give you good examples of the benefits of the investment you have made
- Use open data certificates to indicate that data is being published openly and in a way which follows best practice guidance
- Use common open standards for data and identifiers to make it easier for data users to use the data you provide
- Include clear licences and provide clear guidance on how to attribute data that you provide, so that people who view it on other platforms can understand its provenance
- Provide both data dumps to initialise local databases and form the basis of analyses, and feeds of updates and corrections to enable data users to keep

analyses, and feeds of updates and corrections to enable data users to keep local copies up to date
- Building and maintaining robust, searchable APIs comes with a cost. Consider business models that offer differing levels of support and availability. You may for instance provide an API for exploration only, and/or a fully supported API with expectations of reliability as a commercial service
- Provide APIs that enable tools to provide corrections to the data you make available, and provide clear indications of the process involved in these corrections showing up in that data

User-facing services:

- There are many open data sources that could be integrated into your service – talking to the providers of that data may help you assess its suitability and create a mutually beneficial relationship
- Provide attribution back to the data you integrate into your services. Design interfaces in collaboration with users that help them understand who is responsible for the accuracy of the data you display
- Provide mechanisms to enable people to provide corrections to data. Design interfaces in collaboration with users that help to set their expectations about how these corrections are dealt with and when they might surface.

Intermediaries:

- Ensure that it's possible for the users of your service to trace the provenance of the data you provide back to its original source, so that they can indicate it to their users
- Help drive improvements to the upstream dataset by directing people to contribute and correct the data at source. If this is not possible, provide a mechanism to correct a local version of the dataset and engage with upstream publishers to clarify feedback mechanisms and make them automated if possible
- Directly engage with editing and correcting collaborative datasets. For example, Transport for London are helping improve OpenStreetMap by contributing detailed layouts of tube stations

This prototyping exercise allowed us to think about a number of questions, for which the answers will often depend on the goals, business models and specificities of platforms aiming to integrate local open data in their products. We hope that it helped highlight the existence of a number of quality open datasets waiting to be used, and pointed those aiming use them in the right direction.