# ARITHMUS Collaborative Workshop on Citizen Data and Official Statistics: Report

Workshop Report

Francisca Grommé, Funda Ustek-Spilda, Evelyn Ruppert, Goldsmiths, University of London

15 December 2017


# Table of Contents

# Background

The workshop 'Citizen Data and Official Statistics' was organised by ARITHMUS (Peopling Europe: How data make a people), an ERC funded research project at Goldsmiths, University of London.[1] The idea for a collaborative workshop drawing on citizen science emerged in meetings with the project's advisory group (consisting of representatives of National Statistical Institutes (NSIs) and international statistical organisations). An ARITHMUS Working Paper, *Citizen Data and Official Statistics: Background Document to a Collaborative Workshop* details the ideas and initial concepts that led to and informed the workshop.[2] The overall objective of the workshop was to identify possible design elements of a 'citizen data app', or a platform that could generate data for statistics for government and research. In the following we will use 'app' as a shorthand for a variety of ways digital data can be collected. Although many of the ideas involved mobile phones, other devices or platforms are also possible. The term 'citizen data' refers to data co-produced with citizens and as an expertise and practice between statistical science and citizen science, a concept that is elaborated in the background working paper.[3]

Workshop participants included ARITHMUS researchers; statisticians from national and international organisations; researchers from different universities in the EU; representatives of research organisations engaged in related issues; and information designers and programmers. We started with this configuration of participants to determine if we could develop shared understandings about design principles and translate these into concepts for a citizen data app that addresses public policy topics. If so, then this would set the stage for further work on specifying a concept in relation to potential citizen groups with whom we could begin a collaboration.

The workshop was held on 25 and 26 September 2017 at Canada Water Culture Space, London. It was organized as a series of exercises and activities facilitated by Waag Society, an institute for art, science and technology based in Amsterdam. Participants worked individually or in four small breakout groups and gathered together for plenary presentations, discussions and flash talks about work on related projects. Details of the workshop organisation are outlined in Appendix 1. Collectively this process lead to the identification of design principles, four possible concepts for a citizen data app and milestones for moving from concept to development. The following is a summary of these results.

---

[1] Peopling Europe: How data make a people (ARITHMUS) is an ERC funded project (2014-19; CoG 615588) and involves a team of researchers: Evelyn Ruppert (PI), Baki Cakici, Francisca Grommé, Stephan Scheel, Ville Takala, and Funda Ustek-Spilda.

[2] Grommé, F., F. Ustek-Spilda, E. Ruppert, B. Cakici. 2017. "Citizen Data and Official Statistics: Background Document to a Collaborative Workshop." ARITHMUS Working Paper Series, Paper No. 2. DOI: 10.13140/RG.2.2.18532.88966

[3] We reserve the term 'user' for users of data and statistics and citizens for people who are involved in co-producing data through an app.

# Design Principles for a Citizen Data App

Breakout groups proposed design principles that could potentially be used for a citizen data app. There were some differences but also shared understandings. We have summarised the diverse discussions and ideas into six main categories that are not exhaustive that do not represent a consensus view but consolidate the main points raised by the groups: 1) Civic value, 2) Experience, 3) Privacy, 4) Ownership, 5) Openness and 6) Governance. *Civic value* addresses the question of what public 'good' a citizen data app could serve. *Experience* speaks to how citizens could interact and engage with an app. *Privacy* addresses how privacy and consent could be designed into an app. *Ownership* concerns who could be the owners, curators and stewards of data. *Openness* details the capacity of data to be linked, source code shared, and a design adapted. *Governance* addresses the desired characteristics of an organising consortium, data collaborative or co-op.

## *Civic value*

1. *An app should give something back to citizens*

A citizen data app does not treat citizens as mere data providers or consumers, but as active participants in the design of the app, the production of data and the articulation of problems to be addressed. In return, an app gives something back to citizens that is relevant to them and/or helps them achieve an aim. For instance, an app could have a 'rewards system' that provides incentives for citizens to use an app and share their data. Rewards could range from giving feedback based on a topic that is of interest, to providing alternative ways for thinking or organising their everyday activities. Accordingly, giving back could stimulate citizen participation in the co-production process.

2. *An app should serve a public good*

Serving a public good differentiates a citizen data app from similar apps or platforms. Public good is defined broadly as "for the benefit or well-being of the public" (Oxford Dictionary 2017) to include filling a gap in knowledge for communities/individuals and NSIs, such as providing near real time data and enabling various forms of sharing. A public good can also take the form of addressing a concrete social problem, fulfilling a specific policy objective or strengthening communities.

3. *An app should facilitate civic engagement and the formation of communities around public issues*

By aiming to fulfil a civic good, an app supports the formation of new communities around shared concerns, interests and experiences. By facilitating and fostering civic engagement it could help achieve objectives set by communities and potentially increase trust in public institutions and services (and subsequently trust in the data produced by them). It could thus bring together citizens and NSIs in ways that have not previously been possible.

4. *An app should be inclusive*

A citizen data app encourages and accommodates participation of diverse groups of citizens, for example, people from different age groups, genders, and socio-economic and cultural backgrounds. Such encouragement could occur through consultations, campaigns and training. For inclusivity, an app considers the digital capacities of different groups and addresses this in its design. A citizen data app aims to avoid bias in selecting citizen participants and the data it produces. It avoids stereotyping potentially vulnerable groups in ways that have consequences for policy making and producing 'bubbles' wherein citizens are differentially fed data and insights based on their data profiles.

## *Experience*

5. *An app should be useful for citizens' everyday activities*

A citizen data app ensures that citizens do not have to go out of their way to use it and so it seamlessly integrates into their everyday lives and provides immediate benefits. Although in line with the principle of giving something back, here the emphasis is to create an experience that helps citizens in their everyday activities. For example, a health-related app could provide citizens with information about mobility restrictions that would help them plan their travel and share knowledge about mobility barriers and experiences.

6. *An app should be easy to use and be supported by a strong investment in experience*

A good experience can make or break an app. Even if a citizen data app meets the promise of thinking about official statistics differently, if it does not provide a good experience, it will unlikely sustain the interests of citizens. Therefore, an app ensures a citizen-centric design in the sense of facilitating what citizens, rather than developers, would like to experience when using it. In any case, it is intuitive, easy to use, convenient and time efficient. Consequently, the design of the interface proves at least as important as the functionality of an app. That said, it is also relevant to ensure that an app can run in the background if desired. Moreover, if an app is fun to use, this could improve its uptake by citizens and help promote it. This involves identifying what citizens consider 'fun and easy'. To do so requires including citizens in the design process in partnership with developers, designers and statisticians.

7. *An app should be tailored to citizens' interests and concerns, and facilitate co-production*

A citizen data app allows for personalisation so it is responsive and adaptable to citizen interests and concerns. This could be accomplished by feedback or discussion forums that enable citizen involvement in the ongoing monitoring, adjustment and co-production of categories to suit different interests and needs.

## *Privacy*

*8. An app should ensure privacy by design*

A citizen data app actively seeks to avoid putting (vulnerable) citizens at risk and refrains from generating potentially harmful data for individuals or groups public. This can be accomplished through the secure storage of data and by ensuring that uses of data do not lead to the (re)identification of individuals. These would not only mitigate potential risks to citizens, but could also contribute to building citizen trust in government statistics.

Preferred are designs that enable data to stay on a citizen's mobile device whenever possible and that make different encryption options available. Moreover, citizens should be given the opportunity to choose different configurations of privacy (i.e. different choices or levels of privacy/sharing), and the following should be avoided: aggressive data harvesting, 'creepy' or hidden forms of tracking, and the collection of unnecessary data (i.e., only data which is needed for the proposed and agreed uses are collected).

*9. An app should enable variable and ongoing negotiation of consent*

A citizen data app has a clear moment when consent is requested and given; and transparency is ensured for all types of data collected and the uses to which it will be put (i.e. no 'black boxes'). Rather than consent being given once and for all, citizens should be able to adjust their consent for uses of their data over time.

## *Ownership*

*10. Citizens should be able to access their data archive*

A citizen data app gives citizens access to their historic and current data. This can include app-related data that is only saved on the device, but also the history of interactions with the app. Extractability of data in different formats and its portability (e.g., to different systems) could facilitate citizen engagement and analysis of their data. As such, access could further support citizen participation.

*11. Citizens should be owners of their own data*

A citizen data app ensures citizens are the owners of their own data in the sense that only data they consent to (as discussed above) would be shared. Data that is produced via an app could be co-owned by a data collaborative, an independent body that serves as a 'steward' of data and made up of representatives from universities, national statistics offices, citizen groups, and other public bodies.

## *Openness*

*12. An app should be open to linking with other data sources*

A citizen data app formats data in ways that enable linking with other data sources held by citizens and statisticians. This adds to the value of the data generated by the app but could also add value to existing data sources, if they are linked. Different types of open data could

also supplement and inform the interpretation of app data, such as contextual data (e.g. weather data, transport operational data).

*13. An app should generate open data and code should be open source*

A citizen data app makes both the data and its code open and available. Citizens are given the option to consent to making some or all their data open. When doing so, mechanisms should be in place to ensure anonymity and prevent (re)identification.

*14. An app should specify its relation to existing data sources*

A citizen data app is transparent about its data quality, and its relation to existing data sources such as whether it would complement, enrich or serve as an alternative. Different relations of the data to existing sources used for official statistics could be considered and communicated such as generating new data, providing data that is complementary (enriching) or contextual. This could include citizen interpretations of data about the meaning or causes of patterns in the data. Such forms of 'small' data could thus be combined with the 'big' data generated by the app and enrich understanding and knowledge of phenomena.

*15. An app should be flexible to change in its design*

An app's concept is flexible, that is, adaptable, changeable and alterable depending on not only technological changes, but also the interests and concerns of citizens and statisticians. This calls for an experimental orientation and imagination of the design concept, which should be open to change with changing expectations.

### *Governance*

*16. Organisational accountability*

A citizen data app is governed by organisational accountability. An organising data collaborative, consortium or co-op ensures that the operation of an app is transparent and accountable. This means, among others, that there should be a disclosure policy in place and that it is auditable. Accountability also extends to transparency and responsibility regarding the algorithms that are part of the app's operation.

*17. Sustainable costs*

The design and use of the app is organised in a manner that can be supported long-term, meaning it is cost effective and manageable in terms of work load for NSIs and other members of the collaborative, consortium or coop.

## How to Build a Citizen Data App? Potential Milestones

Here we report on the milestones defined by the workshop participants, which articulate the steps to roll-out a citizen data app to a sufficiently large number of citizens. The list of milestones below does not present a definite series of steps in a pre-determined order. Rather, they highlight important considerations in the development process of a citizen data

app. The milestones can be located at different stages of a development process, yet the processes leading to the milestones generally need to be initiated at an early stage, and many of these processes run parallel to each other.

It should here be noted that, although group discussions generally overlapped during the workshop, a main source of difference was how data generated by an app would relate to existing official statistics. As a result, for apps producing data related to existing statistics different milestones were suggested than for those producing new, complimentary or contextual statistics (see milestone 8).

1. *Internal NSI commitment*

The development of an app needs a NSI champion from high-level management. This could secure, among other things, the dedication of personnel and funding for the development and maintenance of a citizen data app.

2. *Stakeholder support (including citizen consultation and building a consortium)*

Among the stakeholders are citizen advocacy groups, government agencies, public service providers and unions. Different stakeholders can be united in a management *consortium* to oversee the citizen data app. In general, the support of citizen groups and public service providers needs to be secured at the outset. Consultation should identify needs and evaluate and refine the design principles proposed above. Citizen consultation can also identify the risks of data collection for vulnerable groups and to identify groups that would otherwise not be visible in statistics. Support from public service providers and the use of public data may be subject to negotiation.

3. *A business model*

The statistical 'business model' describes the organisation of the processes and systems leading to the production and dissemination of statistics, including aspects such as data collection, quality control, and the adoption of standards. The business model can be developed in a trajectory that parallels prototype development and considers the feasibility of the app for NSIs (especially relevant for apps fully integrated in the statistical production process). If a citizen data app is developed by a consortium of NSIs and stakeholders from different countries, this could include distributing tasks among NSIs and possibly international statistical organisations.[4]

[4] The Generic Statistical Business Process Model (GSBPM) served as an inspiration to some of the milestones in this report. The GSBPM was developed by the statistical office of the United Nations Economic Commission for Europe (UNECE), see: https://ec.europa.eu/eurostat/cros/content/gsbpm-generic-statistical-business-process-model-theme_en.

4. *Funding*

App development does not need large funds from the start. Small funds might be secured for the development of a prototype and a Proof of Concept (PoC). Preparing for roll-out needs larger and sustained funding. Securing funding can be an important role of the management consortium.

5. *App development: Proof of Concept (PoC) and prototype development*

The PoC and prototype are outcomes of a process of co-design/production, most notably with citizen stakeholders. Regarding the development of the technical specifications for a prototype, this could involve the format of a hackathon involving different types of experts. Critically, hackathons should not serve as a way of acquiring free labour. A hackathon would involve informed and paid teams to do goal-oriented work, possibly on an invitation-only basis. An app concerned with the identification and development of new categories and classification could involve machine learning. The PoC and prototype would furthermore need to include considerations regarding privacy, security and data ownership.

The result of this work could be a public API (Application Programming Interface). Next to the technical development of the app, developing the prototype could involve citizens as co-producers of data from an early stage to gather and consolidate a community of developers and stakeholders.

6. *Evaluation, citizen feedback, and adjustment*

Prototyping is a good moment for evaluation, citizen feedback and concept adjustment. That said, frequent evaluation is needed at different stages in realising the app. Moreover, when in operation, feedback could continue to assure an app's flexibility with user needs. However, feedback and flexibility would need to be balanced with the feasibility of adaptation for NSIs.

7. *Pilot*

A pilot would build on the results of prototype testing. The duration could range from six months to several years during which the app could operate as if fully integrated to existing processes of statistical production. Important considerations include setting an end date and establishing the terms of evaluation to prevent pilots from silently expiring. The final stage is to finalise the app and confirm its robustness.

8. *Integration or connection with statistical production process*

This can refer to two outcomes: 1) Data produced by an app is integrated with other data within statistical production processes; 2) Data produced by an app is adopted as complimentary, enriching or contextual to other data within statistical production processes. The latter refers to statistics that do not yet exist but highlight new phenomena, such as citizen experiences and interpretations.

9. *Campaign launch*

A campaign is another way of involving and acknowledging citizens as co-producers. A campaign can start early to inform citizens of the development process and its purposes. It is important to maintain and build trust from the beginning because citizens are being asked to

share their data with government agencies and possibly others (for data that is agreed to be open). Apps producing new statistics would need a more elaborate campaign compared with an app that can be integrated with existing processes for the recruitment of respondents. The tone of the campaign needs to emphasise the importance of citizens in the production of statistics. Moreover, it needs to be similar in tone to communications about the evaluation process. Citizens should have opportunities to provide feedback on the most desirable ways of approaching and engaging them.

*10. Roll-out to a large user group*

A sufficiently large group of citizens using the app and co-producing data is the final outcome of a roadmap with mechanisms in place for ensuring sustainability, credibility and usage.

## Concepts for a Citizen Data App

During the workshop, four different possible concepts for a citizen data app were developed in relation to identified topics and problems. The following is a brief, non-exhaustive sketch of ideas for these concepts (See Appendix 2 for visualisations of these concepts produced at the workshop).

### *How We Move App ('Howie')*

The concept addresses the time and spaces of movement/mobility: the time spent in movement and the spaces of movement. It combines automated data collection (time, location, weather, air pollution counts, transport breakdowns) with citizen provided data on socio-demographics (including mobility needs related to age and health), preferences (for walking, moving through lower pollution areas, step-free access) and reports on problems or conditions affecting movement on specific occasions (elevator/escalator outages). Such data would enable understanding patterns of movement across time and space in relation to factors that help/inhibit movement. It would also show different priorities people have and variations and inequalities in the time/space patterns of movement. Groups could form around shared issues and concerns about mobility, provide advice to each other, and individually and collectively feedback to transport authorities on issues. It would provide more 'small' data and nuanced understandings of how and why people move/are able to move and what their time/space patterns of movement mean.

### *Household Spending App*

The app would be part of the existing NSI data collection processes on household spending and the consumer price index, which include diary-based methods and questionnaires. The new methods of data collection would involve citizens taking quick photos of receipts and products purchased. The product codes on the receipts can then be automatically processed and categorised by the NSIs. In addition, data would be collected through short questionnaires prompted when citizens send a photo, for instance, about how often they purchase a particular product. The app would improve NSIs existing data collection by reducing

respondent burden. Improved data collection, in turn, leads to better insights into household spending and debt. In addition, it could improve data collection on underreported purchases (e.g. recreational drugs) or on purchases otherwise uncaptured in data (e.g. purchases at farmers markets where not all individual transactions are recorded).

To contribute to citizen well-being the app could give feedback based on the interests citizens express during the design process and evaluations. For instance, it could offer to help citizens to lower their spending on groceries by giving advice based on spending patterns of similar households (for instance, the neighbourhoods they shop in). Other options could include giving feedback on plastic consumption, carbon emissions, or health and nutritional value of food items.

### *Civ-O, Making Things Civic App*[5]

Civ-O would be a civic engagement app that brings together various public services and provides an active engagement medium/platform between citizens and providers of public services. When using the app, a citizen can see (depending on their interests) local health providers, soup kitchens or public parks, provide feedback on them, make suggestions for activities and events to public service providers and other citizens, as well as participate in forums about these services. As such, the app would expand what is defined as a "public service" (e.g. a bench could also be a public service if people use it and enjoy using it on a local street). The app would bring together individual citizens and local government service providers in a way that has not been possible before. This engagement could increase trust in public institutions and services (and subsequently trust in the data produced by them). The data produced via the app would not immediately replace existing official statistics such as time-use, public spending or demographic data available at the local level; but it would help enrich them.

### *The Gig Economy App*

The underlying idea is a modular app for building communities around different topics of choice. Tailored to the "gig" economy, the concept would address an emergent phenomenon for which there is a growing need for data: temporary forms of labour that are commonly made possible by digital technologies and apps. This includes for example care workers, freelancers in creative industries, visiting tutors in Higher Education and delivery people. In addition to individual workers, the app would benefit care recipients, patients, unions, service providers, businesses and governments as it could improve understanding of the lived experience of working in the gig economy, contribute to building communities amongst workers, inform public policy in regulating work, and support unions in collective bargaining.

The data collected would include GPS data and would begin and end with the workday. In addition, the app could collect data on measures of stress, different types of work activities

[5] Civ comes from 'civic' and 'o' is used to stress 'our' instead of 'I'.

and those activities which are usually not considered as work but take up time. This data could be compiled via questionnaire type respondent fields. Feedback to users could include mappings of an individual's activities, mechanisms for supporting the formation of groups around shared experiences/interests, and forums for knowledge sharing.

## Reflections on Co-Production and Citizen Data

In this final section, we reflect on the outcomes of the workshop in terms of ideas for engaging with citizens in the co-production of data for official statistics, and issues and questions related to this aim. One of the key questions leading to this workshop is how citizens can actively participate in the production of official statistics. Rather than objects of research, how might they might be active subjects who do more than answer questionnaires or assist with data collection. With this understanding of citizens how then could they actively participate in, for instance, assessing data quality, defining categories and classifications and organising the processes by which this is done?

During the workshop participants developed and articulated different notions of how this conception of the citizen could translate into models of co-production. While these varied in the form and degree of citizen involvement, they identified ways that citizens could be more active. Among these were: consultation with citizens at all stages of development; the inclusion of citizen evaluations in the app development process and as part of the 'flexible' and responsive app design; and on-going participation of citizens in the definition and formulation of new categories and classifications of data. Regarding the latter, one of the benefits and potential gains of co-production identified is how an app could generate new data and categories on phenomena that are usually not visible or understood. Specific examples include how categories of mobility could be rethought considering changing work/leisure patterns or how categories of work could be rethought considering new forms of 'gig labour'. How these results would relate to international classifications and quality standards, which are necessary to maintain credibility, trust and international comparability, is one of many identified challenges of citizen data. This is related to another issue: the geography of a citizen data app such as its uptake in different European cities. To do so would require addressing how it could be integrated into different NSI statistical production systems and adjusted to different combinations of stakeholders, issues and requirements. One possible solution is captured in one of the design principles: that data produced by an app could be adopted as complimentary, enriching of or contextual to other data within statistical production processes. A final challenge highlighted concerns the potential consequences of making citizens more responsible (as individuals or groups) in the production of data for governing purposes. While responsibility might be empowering it could also burden citizens. In general, while the workshop highlighted the positive potentials of a citizen data app, further consideration of potential collateral effects is necessary.

How and under what circumstances citizens can be engaged as more-than-data collectors or respondents in the production of official statistics will be a central question in any follow-up experiment involving the development of a citizen data app. This report has summarized some initial ideas for undertaking such an experiment identified in the workshop which involved statisticians, academic researchers and information designers and programmers. We started

with this configuration of participants to determine if we could develop shared understandings about design principles and translate these into concepts that address public policy topics. The workshop did achieve this and has set the stage for further work on specifying a concept along with potential citizen groups with whom we could begin a collaboration.

# Appendix 1: Workshop Format

Day 1 began with break out groups. In this session, participants discussed apps or internet platforms they would use (or would like to use) as part of their everyday activities and why they chose them. In the same vein, they discussed which apps or platforms they would never use or would refrain from using, and why. These two questions provided some background considerations for groups to identify design principles, which they thought should be part of a potential citizen data app. Groups focused on design principles that could be adaptable and applicable to any potential topic or issue an app might address. This included, for instance, the types of activities enabled by an app, such as networking, sharing, communicating, interactivity, and activity feedback.

Next, the participants built data maps from their perspective as citizens in break out groups. The exercise aimed to draw attention to what kinds of data are being collected about citizens in a typical day through their interaction with various mobile, internet apps and platforms, but also with various government registers (e.g. taxes) or the use of a credit/debit card; border agencies (e.g., some participants crossed borders to attend the workshop) and so on. A striking insight was the sheer amount of data shared and circulated about citizens that it was challenging to identity data that was not shared.

After the data mapping exercise, groups started discussing topics that were of interest to their individual participants towards identifying a group topic or issue they would like to see an app address. This was followed by exercises through which groups sought to translate their initial design principles into a concept for an app/platform. They did this by addressing four questions: “why, what, how and who”. 1) Why would this topic be a good topic for a citizen data app? 2) What would the app involve and what kind of gaps in data (or policy needs) would the app address? 3) How would the app be different than all the existing apps out there and how would it be useful for official statistics and 4) Who would be the target audience of this app, who would be its “users”?

Day II of the workshop built on the first day discussions and involved groups further specifying their concepts and designing interface screens with which citizens would engage. To focus their apps, groups were asked to address the *urgency* of and necessary *supports* to their designs. Urgency meant that groups consider who would desire/want such an app? Who would be disadvantaged without it, and who would benefit the most from it? As such, groups were asked to think about who would be most likely positively affected by the outcomes of the successful development of their app. Similarly, groups were requested to hypothesise on who would be the different stakeholders involved in building their app; and how they could become invested in the process and data product. In short, groups were asked to think about what would be needed to champion the design principles for their app.

Moreover, groups extended their discussions to add further layers of detail to their concepts by elaborating on the following questions:

1. Citizen: which citizen organisations and groups need to be involved? Where can you find them (on/offline)? What is the best moment in the process to involve them? How can you ensure inclusivity?
2. Process: how might the app integrate with existing data collection and statistical processes? How could stakeholders be updated on its progress? Who would be

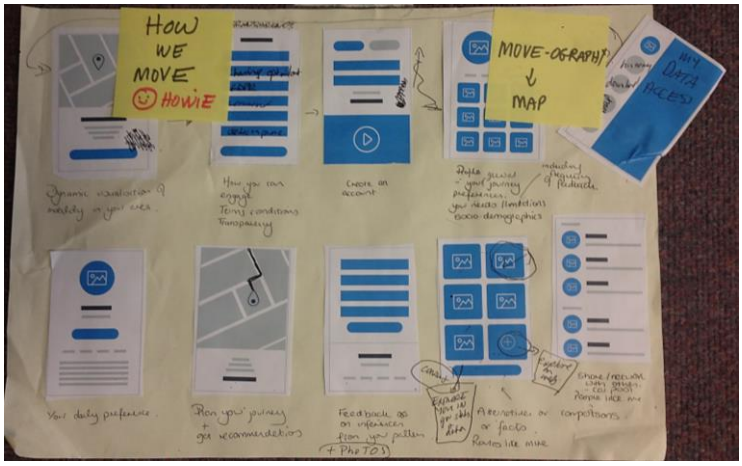
responsible and at what moments? Which (new) roles and decision making moments are needed?

3. Campaign: which communication means would be best for promoting the app? Which channels are currently at your disposal? What would be the 'tone of voice' for promoting the app?
4. Ownership: what would be the possible models of data and app/platform ownership? How could ongoing maintenance and updates be organised?

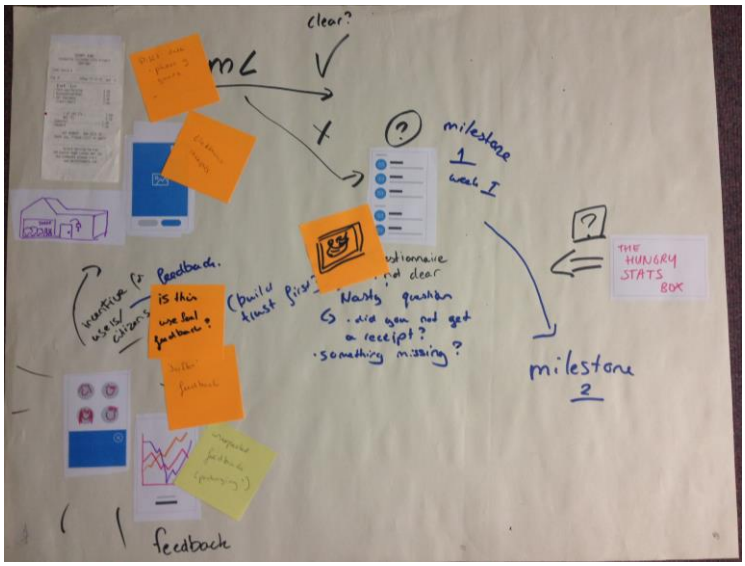
At the end of the day, all groups presented their final concepts, roadmaps and milestones for building their proposed apps. Each presentation was followed by a plenary discussion on the main ideas and issues arising from the concepts and designs.

# Appendix 2: Sketches of App Concepts

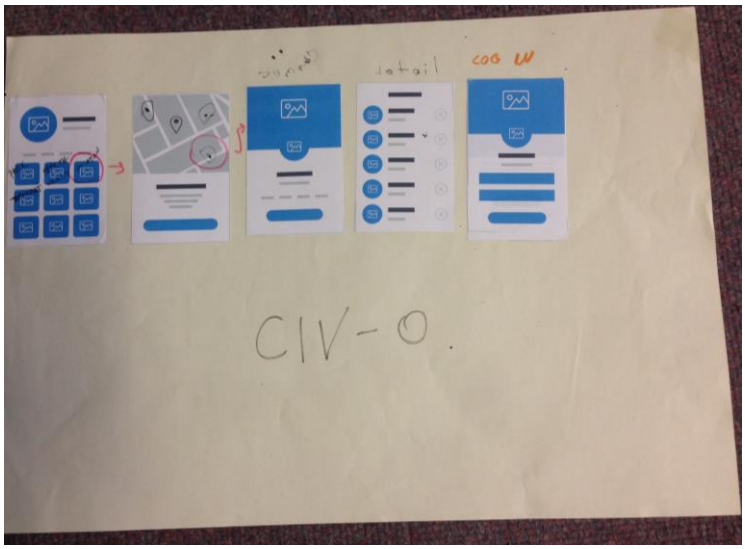
### How We Move App ('Howie')




### Household Spending App



## Civ-O, Making Things Civic App




## The Gig Economy App