Privacy Measures for The Big Data Show

Version 0.7 28th August 2020

v0.7
28th August 2020

Contents

v0.7
28th August 2020

1. Introduction

Due to the Covid-19 pandemic the plans for The Big Data Show were cancelled in March 2020 This document is an update to v.05 which was written for the live show. It describes the project as it is now online.

This document has been prepared and approved by the Ethics Committee, taking on advice and guidance from key partners including Police Scotland. It presents the privacy and data protection measures adopted by *The Big Data Show* (TBDS).

TBDS will be delivered through a website https://thebigdatashow.online ("the website") this has been developed and designed by Foxdog studio and will also involve the use of a smartphone app ("the app"), called *Super Swipe.* This is being designed and implemented with the help of Two Tails Studio, a specialist in games and software development. Documents outline the security measures for the app and the website are attached to this document in the appendix.

Up to 30 shows will be presented in Perth, Glasgow, Edinburgh through the website and app. In-school workshops will be delivered to each participating school when it is safe to do so.

The producers of TBDS aim to improve the data literacy and cyber resilience of pupils who take part in the project and particularly to inspire young women and people of colour to consider careers in digital technologies.

Outcomes for: children, teachers, education system, government, Civic Digits

Short-term outcomes (immediate):	**Medium-term outcomes (eg after six months):**	**Long-term outcomes (after a year):**
Awareness of cyber issues is raised Ethical hacking careers are better understood Learners achieve SCQF credit Teachers feel supported to deliver curriculum Curriculum delivery is recognised as being creatively enhanced Funders' requirement are fulfilled	Cyber awareness is raised Children have progressed onto further cyber-related/digital or STEM learning opportunities as a result of TBDS Interest in cyber/digital/STEM careers is increased A model of creative delivery of STEM learning is available, with evaluation Teachers have explored more opportunities to bring creative digital/STEM learning into the curriculum (eg cross-departmentally within the school) Scotland's progressive approach in creatively teaching cyber/digital awareness is recognised	This approach is common and growing in Scotland's schools Children are on the path to further cyber/digital/STEM learning The model has run successfully outwith Scotland We have concrete evidence of an increase in learning and achievement and aspiration from the children in this area We have evidence of increased capacity in schools and between schools and other organisations to deliver curriculum in a creative way.

This risk assessment has been written by Dr Clare Duffy, Artistic Director of Civic Digits with Rupert Goodwins, co-writer of *The Big Data Show* and Creative Technology Officer of Civic Digits

This is a dynamic document and will be reassessed on at least a bi-annual basis.

2. Organisational Structures

2.1. TBDS Ethics Committee and Education Steering Committee

Ethics Committee

Person	Affiliation	Role
(EK) Prof. Ewan Klein (Chair)	University of Edinburgh	Use, interpretation and social embedding of different forms of data advisor
(NC) Dr. Natalie Coull	Abertay University	Ethical Hacking Skills Advisor
(LP) Dr. Lynn Parker	Abertay University	Games development advisor
(DS) Daniel Sellers	Scottish Government:	Cyber Resilience Learning and Skills Policy Adviser
(FO'B) Freda O'Byrne	Independent Artist Co-founder Prewired	Theatre In Education advisor
(AL) Allan Lindsay (RM) Rachel McKay	Young Scot	Participation and Co- design Director Co-design manager

Education Steering Committee

The Education Steering committee are convened to help us achieve our long term aim to embed TBDS into the education fabric of Scotland and the UK and to advise on the educational outcomes of the project. They are

Caroline Donald (Chair)	Head of learning and Engagement	Edinburgh International Festival
Kirsty McFaul,	Senior Officer for Digital	Education Scotland
Scot Hunter	Internet Safety Officer	Education Scotland
Debbie Bentley	Drama Teacher	Firhill School Edinburgh

2.2. Civic Digits Team

Person	Role

(CD) Clare Duffy	Creative Director Civic Digits
(SG) Suzy Glass	Executive Producer
(RG) Rupert Goodwins (RJB) Robyn Jancovitch-Brown	Creative Technical Officer Project Manager

2.3. Two Tails

Director David Mitchell: Delivery of the app for use in live performance and workshops.

3. Stakeholders

3.1. Internal Stakeholders

Co-Producers

- Civic Digits C.I.C
- Perth Theatre
- Unlimited Theatre

Partners

- Edinburgh International Festival (Creative Learning)**,**
- The Citizens' Theatre (Glasgow, Creative Learning)
- The Lyceum Theatre, (Edinburgh)
- TBDS Ethics Committee (Details given above)
- The Educational Steering Committee (Details given above)

3.2. External Stakeholders

- Schools: 30 across Edinburgh, Glasgow and Perth.
- Education Scotland
- Funders: Creative Scotland, Scottish Government. Cyber Resilience Unit, Digital Xtra, Garfield Weston
- Creative Informatics. The University Edinburgh.
- Business sponsors: Skyscanner
- Advisers: police, press contacts,
- Ciara Mitchell ScotlandIS: Head of Cyber

4. Policy

4.1. Child Protection

Clare Duffy is a registered STEM Ambassadors and therefore has an up to date PVG Disclosure. Other members of the creative team such as actors, co-writers and other creatives do not require PVG disclosure for accompanied, one off visits to schools.

4.2. Privacy Statement

Civic Digits has a Privacy Statement which conforms with the requirements of GDPR. Civic Digits will hold data about teachers, schools and public audience members who consent to it but will not share any of this personal data with any third parties without consent.

Civic Digits will share anonymised data about audience and participant numbers, age groups, and geographical location of performances for research and evaluation purposes, for example with funding bodies such as Creative Scotland and with academic institutions.

The privacy statement is available on a website before the workshops begin. 'Plain language' Privacy Statements will also be given out to audiences at the end of each show to make it clear that no personal data has been seen or collected by Civic Digits though the process of delivering the show.

4.3. Terms of Service for the app

Civic Digits have written a statement of Terms of Service for using the app, with the guidance and final approval of the Ethics Committee. The aim of the Terms of Service is to replicate as much as possible the typical experience of a user downloading and using an app from Google Play or the Apple App Store. We also want to use this document to show how the language and form of these terms can be very hard to understand by including absurd and/or silly statements. These absurd statements will be referred to in the workshop and there will be further suggestions of ways to explore them as part of our challenge to the users to think critically about their relationship with the makers of apps.

4.4. IP and Licensing

The intellectual property for the app is shared equally between Civic Digits C.I.C and RG. Civic Digits C.I.C and RG propose to release any source code for the app that does not already fall under previous licences under the GPL4 Licence. This source code will either be available via the website or by email request.

Confidentiality

All participants will be informally asked to agree not to disclose the way this project and performance works. Much of the delight and magic of the experience is based on testing what the participants do in a typical online situation. Alerting the audience to 'the game' would genuinely spoil the experience. However, we will not require participants to sign a non-disclosure agreement.

4.5. Project Evaluation

Quantitative evaluation.

The project will offer 30 pupils per school the opportunity to achieve the SCQF-credit rated award certificates. This data will be managed by the participating schools. Civic Digits will have a code that relates to a school register and will have no access to pupils' personal data.

This 'Introduction to Digital Citizenship' course will involve participation in 2 workshops and will also require outside of direct contact with Civic Digits the sharing of creative work using data with the rest of the school and gathering feedback.

Qualitative evaluation

The production team delivering the shows will create a 'show report' for every show in the tour. This will reflect the whole team's assessment of how the participants responded to each element of the workshops. This will not include any personal data from participants.

The workshop leaders will assess and award the SCQF certificates in the second workshop and will record details about how learners experienced the project.

4.6. Social Media

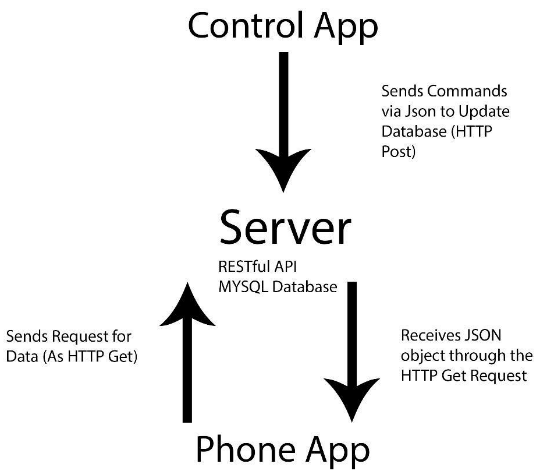
TBDS has a Twitter account, namely @bigdatashowtime. It is likely that Civic Digits and TBDS will create a variety of social media profiles in the future. Social media platforms sell, use and/or share data which can identify individuals and threaten their privacy and security for financial and potentially other reasons. Civic Digits' aim is that the work we do will contribute to the debate about the role and responsibility of us as citizens and the organisations who offer these platforms to know about how they use (personal) data and how that affects privacy and security online and offline.

4.7. Customer Relationships Management (CRM) of Data for schools

Civic Digits will keep a CRM-style database of the schools and teachers it has contact with to maintain communication throughout the project. In some cases, this will include personal telephone numbers for teachers. Consent to hold this data will be requested with booking instructions.

5. Functionality and Data Handling of the app

5.1. System Data Flow


Control App
Sends Commands via Json to Update Database (HTTP Post)
Server
RESTful API
MYSQL Database
Sends Request for Data (As HTTP Get)
Receives JSON object through the HTTP Get Request
Phone App

An overview of the data flow is shown in the accompanying diagram. In this diagram, the app described elsewhere in this document is referred to as "the Phone App".

The Control App does not ever interact directly with the Phone App, but sends data to the Server. The Server stores this data in a database. During a TBDS performance, the Phone App connects regularly with the Control Application to "ask" it what scene it should be displaying, whether it should be playing a sound effect or triggering any other "magic tricks". The app also sends a game score to the server and receives a high score table back, when the phone is known to the servier by a long, random key that exists only for the duration of the show and for this single purpose: nothing else about the phone or its user is known to the server software. Data extracted from the database by the Server is sent to the Phone App as a text object (JSON).

RG as CTO wrote a security design specification. In the appendix to this document Two Tails have answered each point of this specification. RG will also have eyes on the code for the app.

5.2. App Permissions

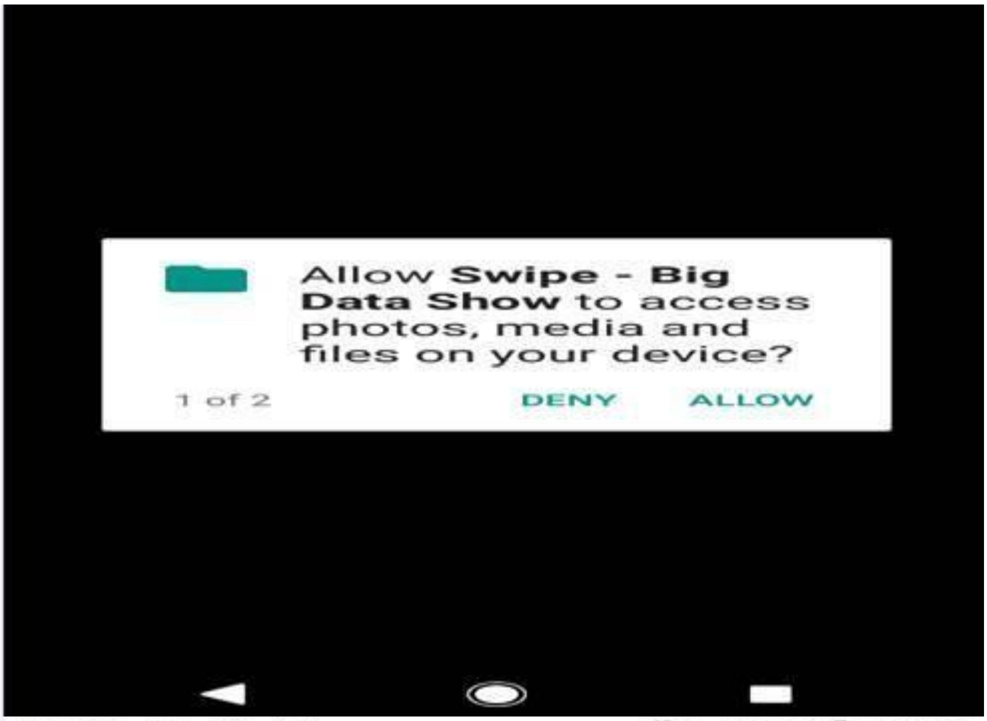
The app will ask the user to allow access to data such as GPS. However, the app will not in fact be able to collect this data either, in the core code or via Android data access permissions as the functionality required to collect this data will not be programmed into the app. According to the Android documentation,[1] these are permissions

> *to access sensitive user data (such as contacts and SMS), as well as certain system features (such as camera and internet). Depending on the feature, the [Android] system might grant the permission automatically or might prompt the user to approve the request.*

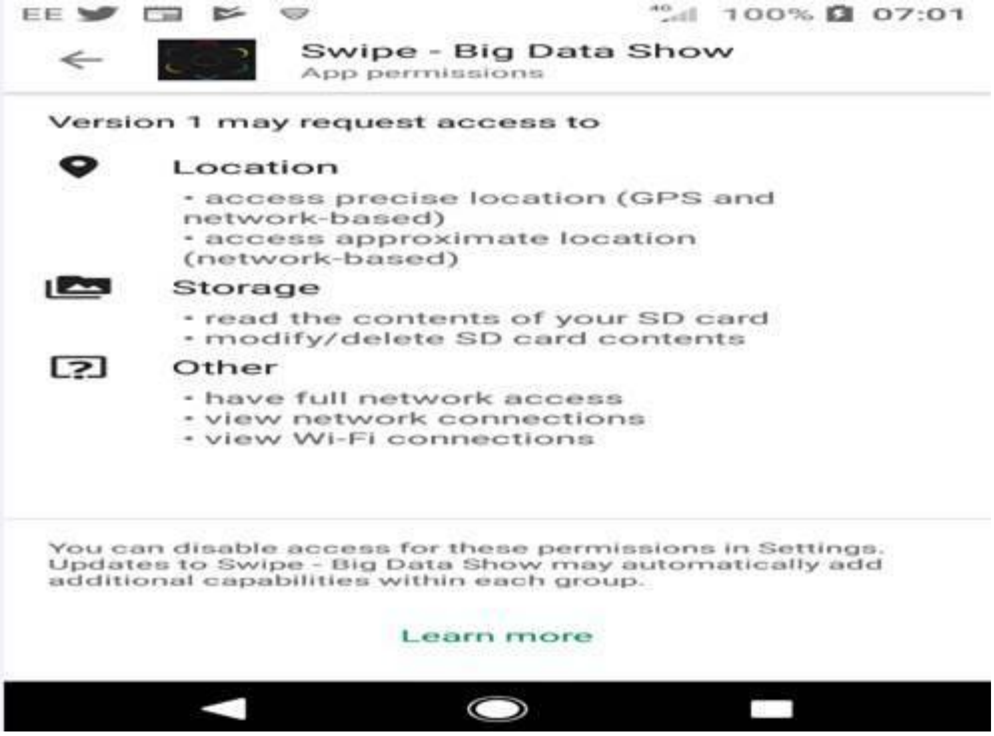
We consider it unethical to allow our app to have the functionality to request data access in this way, since if there was a malicious act on the phone, the data could be exploited.

When first downloaded, the app asks for permissions to access storage and network connections.

[1] https://developer.android.com/guide/topics/permissions/overview



Allow **Swipe - Big Data Show** to access photos, media and files on your device?



Swipe - Big Data Show

App permissions

Version 1 may request access to

Location

- access precise location (GPS and network-based)
- access approximate location (network-based)

Storage

- read the contents of your SD card
- modify/delete SD card contents

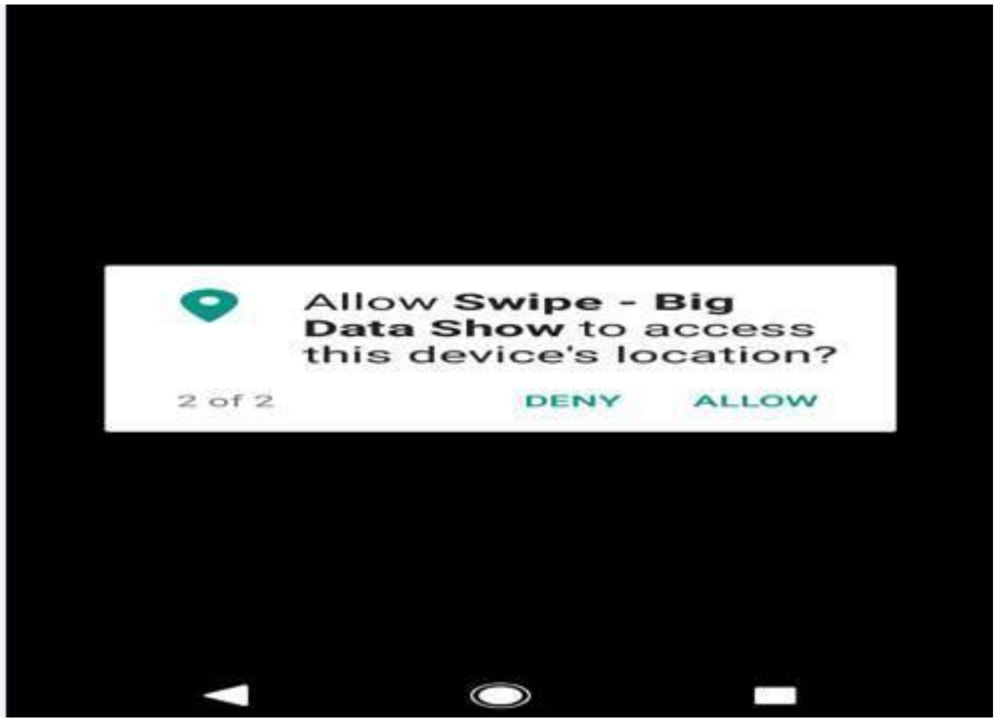
Other

- have full network access
- view network connections
- view Wi-Fi connections

You can disable access for these permissions in Settings. Updates to Swipe - Big Data Show may automatically add additional capabilities within each group.

This is a standard for Apple and Google Play and beyond our ability to change. However, our app will not access these areas. If we made this clear before the show we believe this would destroy the potential learning impact of the experience. For this reason we will distribute a simple and 'Plain English' Privacy policy/declaration to all pupils after the show.

5.3. App Data Collection



Allow **Swipe - Big Data Show** to access this device's location?

No user data is transferred from the user’s phone via the app except a single high score figure known by a long, random, temporary key shared between the phone app and the server. . The only information that will be recorded from the user in the app is their GPS location. This data will be stored locally (within the user app file structure on the user's phone) and never leaves the phone. This GPS data will be deleted as soon as it has been used in the show.

For players of SUPER SWIPE who do not attend the show, the app will be updated at the end of each tour. Every update or re-installation automatically deletes previously collected data.

All analytic data recording has been disabled for the app itself. Some analytic data may be recorded by the Operating System (Android or iOS) or by the devices themselves (for example the Apple App Store may record when the application was downloaded) but this is standard for all phone apps, is outside of our control and is GDPR compliant.

There will be a competition at the start of the show to find ‘The Swipe Champion’. This will be announced by a message sent to the highest scorer and will not require any personal data to be seen or collected.

5.4 Live looking over the shoulder of other audience members

There is the risk that an audience member can see data ‘over the shoulder’ of a class mate when the GPS data of the user is presented to them in the form of a map of where the user was when they played SWIPE. The map is specifically designed to show the user where they were, but to make it difficult for a glance to reveal very much information. Each point on the map is visited for only a few seconds and there are no place identifies such as street names or organisations addresses (for example an LGBT youth club). The resolution of the map is not particularly fine grained, each view of the map representing approximately 0.25 km.

Privacy Risk Register

Note: Risks are ranked on a 1—5 scale. Exposure is an estimate of the combined effect of Impact and Likelihood.

Risk description	Inherent Privacy Risk			Options for avoiding or mitigating t
	Impact	Likelihood	Exposure	
Reputational risk to Civic Digits if there is a perceived risk (e.g., via press or social media) of privacy invasion.	**Mild (2)**	**Mild (2)**	**Mild (2)**	Ensure demonstrable adherence with data prote Ask a highly qualified ethics committee to suppor think through all of the risks.
Spoof versions of the app are installed on the Apple and/or Android app stores and allow malicious code to be installed on users' phones.	**Low (1)**	**Low (1)**	**Low (1)**	Ensure clear branding and make it as fool proof a participants to only download the correct app On a par with any other app in the ecosystem. It impact as a spoof app could do anything it liked, no more risk here than do other apps.
Faulty implementation of the app allows personal data to be transferred to the system Server.	**Low (1)**	**Low (1)**	**Low (1)**	**Server is protected by strong passwords, and is c accessible to authorised members of the team.** TBDS does not collect any personal data on the p GPS tracking. The architecture of the system, the server and the functionality of the back-end syste store any data of any kind during operation, so it pathway which could malfunction - there is no pa
Insecurities with and on the server allow third parties to tamper with the Server routines to send malicious payloads to the user app.	**Low (1)**	**Low (1)**	**Low (1)**	**The server is a standard regularly patched currentl instance hosted on AWS and managed by a reputa with whitelisted 2FA admin access to the hosting s to the instance itself.**
Insecure data transfer between user app and Server is breached and malicious payloads are transferred to the user's phone.	**Mild (2)**	**Mild (2)**	**Mild (2)**	No user data is transferred between the server a and there is no pathway for malicious data onto t risk that an unknown vulnerability in the app cou custom-designed attack, but that's true for all ap expected audience for the app is small, compare the audience is mostly school children with little impact. It is considered unlikely that it would rep required to attack it.
Malicious code on user's phone is able to access private data stored on the app's local file store.	**Low (1)**	**Low (1)**	**Low (1)**	The only private data stored is recent GPS trackir exploit in any harmful fashion for our audience, a little motivation to any malicious attacker to inve extracting it.

Malicious software on one user's phone attacking another phone connected to the server	**Mild (2)**	**Mild (2)**	**Mild (2)**	All phone traffic to and from and server inputs are bounds see each other's traffic, and t successful malicious attack th the server is considered low c

				the installation and the use o precautions. Server instances with current patches.

Appendix

A: Security measures for the App “Super Swipe” and its control system

1. All communications controlled or specified by TBDS are encrypted. This includes game-server, control_app-server, and game-mapping_service.

   All communication between app, server and control app uses HTTPS/TLS encryption.

2. All communications protocols between game-server and control_app-server are explicitly documented by Two Tails and disclosed to TBDS. These include which APIs are being used, data structures, message sizes, timing, and transaction descriptions.

3. API documentation can be found here: https://docs.google.com/document/d/1QIQdy5dTq8VXjrN5dxCu2k9t605LcUISY5p1Sh3LPXs/edit

4. APIs implement suitable security practices to ensure user access is restricted where necessary, for example, e.g by using API keys.

   Write endpoints (POST/PUT/PATCH) used by the control app require a valid API key to access. Requires with an invalid API key are rejected.. API keys are managed by a server admin. API keys are only valid for a certain period of time before automatically expiring. Expired API keys cannot be reused.

5. Geo mapping service API and throughput requirements

   Mapbox free tier is available for up to 25,000 monthly active users, which gives plenty of room for TBDS. https://www.mapbox.com/pricing/

6. Overall network traffic per user, average and peak, and an indication of server CPU/memory requirements for 1.5K users.

   Outside the show, the app makes no requests to the server. During a show, on average the app makes a request to the server every 5 seconds. This request is typically around 800 bytes including request and response.

   Therefore over a 10 minute period, one user would typically use around 93.75KB of bandwidth (includes in and out). 1.5k users would use 137.32MB.

   The server software itself is very lightweight, making use of a single PHP instance and a MySQL database. A small virtual hosted server with 2G of RAM and a standard CPU configuration will be more than capable of handling the requirements for the show.

7. All data structure handling is bounds-checked and explicit handling provided for malformed or out-of-bounds messages. Any malformed requests to the server are rejected.

8. Control app, game authentication, DOS attack mitigation

   DOS protection is difficult as there is no single easy solution and a proper solution requires an experienced server administrator to put in place and monitor.

9. All code for the app, control app and server is hosted securely on Bitbucket and access is granted to members of Civic Digits.

B. The Big Data Show Website Security and Privacy

17th Jul 2020

- The server is a virtual machine provided by Linode running Arch Linux. Only the directors of Foxdog have access to the Linode account and their passwords are stored by Lastpass.

- The virtual machine is dedicated to the site and is not used for any other purpose.

- Access to the server is restricted to SSH with key-based authentication. Password authentication and root login are disabled.
- Administrative access on the server is protected by sudo with password authentications.
- The server is updated weekly to apply security and stability fixes.
- All communication with the server is protected by TLS. Our certificate is provided by Letsencrypt and the web server's TLS layer is configured by certbot.
- Streaming to the server is protected by a password, which is only shared with the operator.
- The remote control interface is protected by a password, which is only shared with the operator.
- The website's admin panel is password protected by Meteor's authentication system. The password is only shared with operators of the site.
- An audit of client-server communication points was conducted to ensure each has the appropriate authentication checks. No modifying (e.g., changing show name) operation can be performed without authentication.
- For non-administrative use of the site, no identifying information is stored on the client or server.
- All information entered by administrators (show names, portraits, etc.) should be considered public and contain no identifying information.
- Removing a show also removes all uploaded images associated with that show.