CoBox
in context

An overview of data management concerns
in a group of co-operative organisations

Sophea Lerner
Jaya Klara Brekke
Mooness Davarian

for CoBox - a Magma Collective project

twitter: @CoBoxCoop
url: https://cobox.cloud

This project has received funding from the European Union’s Horizon 2020 research and innovation programme within the framework of the LEDGER Project funded under grant agreement No.825268. This report reflects only the authors' views, and neither the Contractor nor the EU is responsible for any use that may be made of the information it contains.

Table of Contents

Summary of Recommendations

Technical Design and Implementation

- **Implement blind replication as a priority for all mutual back-ups.**
- **Ensure that no single node can accidentally become responsible for destroying or exposing sensitive data belonging to someone else. Data-hosts must not be unreasonably liable for their peers' data, which they store.**
- **Critical functions, particularly those affecting other people's back-ups and key-shards, need to have robust default settings that can only be set to 'manual' in-extremis. For example, if synchronisation needs to be temporarily halted on a given network, the software should strongly prompt for when it should resume (similar to putting your phone on silent) rather than just switching off (as with Nextcloud sync).**
- **Allow an option for decisions to be logged with information about rationale at points where change of personnel could lead to breakage or lack of context could slow down recovery after a failure.**
- **Manage transfers of larger back-ups without crippling network speeds.**
- **Allow users to allocate bandwidth around their working hours.**
- **Work should be possible online or offline.**
- **Multiple users should be able to work on a file with varying connectivity without causing version conflicts.**
- **Encryption on top of CoBox should be seamless.**
- **High granularity in sharing permissions for the complex range of stakeholders these coops often engage with both internally and externally.**
- **Fast and safe ways to back up user keys are critical. Given the challenges many people face in choosing and managing secure passwords this is the area that could make or break CoBox.**
- **Ability to transfer and share data within an organisation without using corporate cloud services, reducing dependence on internet connectivity. If this is achieved by substituting with a local network protocol, care should be taken to replace the default off-site backup function that mainstream cloud services perform.**
- **Build in support for non-technical professional tools that do not force an all-in approach with locked-in data.**
- **Data retrievability is an extremely high design priority, even if everything else breaks.**

- **Design problems should be solved in standards-based ways to future-proof the work involved for users onboarding with CoBox. This may require some aspects of team's workflow to judiciously marry a 'slow and careful' requirement with the benefits of rapid iterative prototyping,**
- **Ease of retrieval in case of local loss. Local data loss is a cause of stress or the result of an otherwise stressful situation. Low stress approaches to retrieving data need to be considered for a number of scenarios. Even where data cannot be instantaneously reconstituted, users should be made aware of the steps involved and expected timeline, so that overall pain is kept to a minimum.**
- **Users felt most mastery over the tools they were using when things worked smoothly, and worried most when they did not understand what was going on. Design decisions should unintrusively support awareness of things working well, of users being in control of their infrastructure, and not exclusively responsible for that of others. Ie. It should be easy to see that both your data and the data you host is replicated in enough other places to be safe if your building floods.**
- **If data-hosts need specific legal protections then the details of that need to be considered from early in the design process.**
- **Many of the target group work in high trust environments with shared access or role-based rather than individual access. Roles may include multiple individuals who all access an account or may rotate between several individuals over time. CoBox credentials and key-management need to work with, and support this reality. 2FA that could be easily switched between different group members personal phones or emails or sent to a group of contacts would be an example.**
- **Provide easy ways to navigate options to mix and match standards-based tools that might work on top of CoBox.**
- **Just as a successful car sharing business needs to make using their service both cheaper and less hassle than owning a car, CoBox needs to offer some of the benefits of running your own server without the hassles.**

Business Strategy

- **Factor key management, security updates and tech support planning into the service as a whole rather than e.g. treat support as an independent function.**
- **Find an equitable system for balancing resources between users with large differences in the volume of data they are sharing. Internet speed, electricity consumption, hardware maintenance etc. need to be considered.**

- **Plan to have a sustainable free option, or non-monetary option. e.g. users could receive credits for providing support, for how much data they host or for providing bandwidth.**

- **Partner with existing providers for roll out and ongoing support functions where they already have niche expertise which the CoBox team may not currently be set up to offer.**

- **Roll out would work best if CoBox comes in alongside other tools and does not try to take over all the corporate cloud functions in one go.**

Ecosystem Support

- **Support cooperatives in clarifying their own transparency and data consent models, so that these can be made explicit to stakeholders.**

- **Include help with password management basics as an option in roll out and training.**

- **As a project that may intimately connect organisations via technology in ways that depend on human networks, CoBox has an opportunity to broker a sharing of practices either as a planned dimension of support and roll-out or as an independent resource developed from this research.**

- **Contingency for CoBox end of life. Design choices should account for what will happen to ensure no-one loses their data if CoBox ceases to operate for any reason (the coop disbands, the product turns out not to be viable, the product is a great success but becomes obsolete because of unexpected new technology, etc.).**

- **During the testing phase it would be helpful to run a parallel back-up with proven technology, think of it as training wheels in-case CoBox falls over.**

- **Support for setting up and maintaining CoBox needs each organisation to have a paladin for the project who is supported according to their need by consistent personnel.**

Documentation

- **Be transparent about what kinds of security threats CoBox is, and is not, immune to, and under what conditions.**

- **Work with prospective users to understand the existing complexity and opacity of the trust models they currently employ as way to provide a context for CoBox**

- **Threat modelling, with a user-friendly interpretation, to help users make informed decisions when storing sensitive data or providing information to clients on how their data is stored.**

Introduction

The CoBox team set out to research data management practices and needs of a selection of British co-ops who are broadly representative of a potential user base for the project. The results reveal the ways in which this sector forms a complex ecosystem with a set of common concerns and specific challenges that CoBox can address.

Methodology

Interviews of varying duration were conducted with 10 participants from 9 co-ops. Interviewees and co-ops were selected for a diversity of technical perspectives and skill levels/types rather than as a homogeneous target user group. The interviews varied in style between interviewers and in one case responses were reconstructed in summary after some audio was lost.

The conversations discussed both general operational data management and software choices, as well as practices relating to each co-op's domain-specific expertise or unique business situation. Interviewees themselves can be divided into four broad groups: i) professional systems administrators, ii) generally tech-savvy 'de-facto sysadmins' who had stepped up to organise their co-op's operational data or engage external tech support in general or specific ways, iii) individuals whose day-to-day role involves building and implementing technological solutions in areas other than systems administration and iv) people with varying tech skills whose roles required them to interact with systems set up by others.

The interviews reflected the embedded involvement of CoBox team members with their target user group, and these personal connections were also fruitful from a relationship-building perspective that will be an important component of a successful CoBox implementation. In many cases, some of these exploratory questions also provided value to participants, offering a useful opportunity to reflect on their day to day practices. Interviewees were generous with their own insights into the wider software landscape that CoBox sits within, and demonstrated a lot of goodwill towards the project that suggests many of them will continue to be supportive and give feedback throughout the process.

Data Retention

Source interviews for this report have been restricted to encrypted storage devices with a policy of restricting access, and destroying working copies when not actively in use. After reporting is complete, audio recordings will be digitally shredded and only anonymised transcripts securely retained for a finite period as a resource to inform the CoBox design process.

Outline of Findings

The open and varied style of this data does not encourage a precise comparison of 'apples with apples' on an issue by issue basis, or highly structured enumerable responses to a narrowly defined set of product-oriented questions. Rather, it provides insight into a highly interrelated ecosystem that CoBox aims to contribute to. The extent to which it clarifies certain themes occurring across different contexts, and reveals varying perspectives on a set of common challenges, has implications for where CoBox will direct its focus and recommends a peer-driven design process.

Overall, the research suggests that there is potential for CoBox to address specific needs relating to mutual data back-up in ways that satisfy a widely expressed desire to be more independent of 'big data' companies, if it can also make the shared responsibility low-stress and offer tangible advantages over aspects of both self-managed independent infrastructure and the corporate cloud in terms of cost and effort.

By working within the existing ecosystem and making current inter-dependencies apparent, CoBox can potentially build a layer of resilience into the current network. In addition to more traditional software design considerations, such as easily being able to see the state of back-ups, participants placed a high value on dealing personally with known individuals. A successful approach to the design challenges would need to include a critical people-based requirement, building on existing relationships to plan support for uptake and longevity. Moreover, it became apparent that the successful switch from the corporate cloud to open source alternatives will generally depend on key individuals being able to drive this process within an organisation.

Data management needs

The main operational types of data itemised by interviewees were: external contacts (clients / customers / contributors), email, chat and other communications data, administrative accounts and passwords, newsletter subscriptions, time tracking accounts and invoicing. Domain-specific data included large real time media, digital art projects, accounts files, web scraping data, print production files, crop data and document archives, student info and attendance, confidential volunteer data, historical data for technical processes, subscription address info, shared access text for editorial processes, residential and visitor data to be shared with municipalities, and client data in various forms (email, hosted websites etc).

Data Protection

Regulatory considerations for co-ops and their data which were specifically identified included various regulations associated with being an education provider or managing sensitive data about volunteers, GDPR, Taxation records, Transaction info to share with municipalities / tourist tax data and copyright.

Responses to questions about data in general were often skewed towards 'other peoples'' personal data in particular and several participants expressed fear related to handling personal data and GDPR. Increasing confidence that "Data" is safe may require a mixture of technical, interface and human elements. Adding more general terms to the discussion during peer-design sessions might help broaden the scope of data under consideration. Negative feelings about GDPR were focused on compliance, fines and losing mailing list contacts. Constructive expressions about it tended to be focused on providing informed consent and safeguarding others.

Outsourcing certain kinds of data handling to external specialist services was a source of reassurance to some that the problematic 'Data', such as payment information, would be cared for by someone else better equipped to do so. GDPR was cited in relation to thinking about backing up other peoples data as a reason why they would only want to host totally encrypted data that they could never access. This points to blind replication as a priority for all mutual back-ups, and support for coops in clarifying their sometimes vague relationships to their own transparency and data consent models, so that these can be made explicit to stakeholders.

A successful implementation of CoBox would need to actively show that it can mitigate this perceived problem in a way that no single node can accidentally become responsible for either destroying or exposing sensitive data belonging to someone else. It would also need to be transparent about what kinds of security threats it is, and is not, immune to, and under what conditions, in order to allow informed decisions about where CoBox fits into the data landscape of a particular group. Related considerations include key management, security updates and tech support planning.

Data Resilience, or Back-Up

Information about specific data *volumes* that people handle and need to back-up was not gathered. Volume can vary tremendously depending on the main business of the coop. For example, the team spoke to people working with high volume dynamic data online, high volume static archives, locally handled production and layout files, as well as people whose work involved handling client data of unpredictable sizes. Those with high volume domain-specific data unsurprisingly already had local or specialised back-up strategies in place. Operational data back-up was more variegated.

For several groups back-up and file-sharing were both achieved via use of cloud services (inherently off-site) with in-built versioning. The principle of 'distribution as an archival method', put forward by one interviewee, could to some extent also describe how a lot of operational data is backed up by default: By being on several users' computers and in the cloud in a fairly ad-hoc way. It is often also not the case that this data can neatly divide into static/archival and current/dynamic. Interviewers did not go into specific file management schema/folder structures in depth in most cases but some people did mention working in search-based ways while others referred to having more structured systems.

Data back-up and redundancy were already handled by third parties in many cases. This was generally the easiest way to ensure that all back-ups were not in the same physical location. Only one respondent described a back-up system wherein they were responsible for manually moving an off-site back-up of a large data-set on a regular basis. In practice this did not always happen. In most cases where back-up relies on manual processes these are often performed irregularly.

Existing back-up strategies in the case of data-heavy and tech-oriented coops included RAID systems, maintained both internally or externally, and back-up and archiving software such as Borg and Bart for making encrypted back-ups.

Stories of data loss and resources involved in retrieval were not requested across the board. However the role of having a record of decisions made, and not just an automated deployment, in getting crashed systems back-up was mentioned as a useful strategy in situations of shared care for a system.

CoBox will need to figure out how to manage transfers of larger back-ups without crippling network speeds by default and to make large differences in the size of different group's back-ups play out equitably in the network with regards to Internet speed, electricity consumption, hardware maintenance etc. A focus on operational data and on supplying an additional layer to parallel current back-ups may provide some leeway with these challenges in the prototype stages.

File-Sharing

The two main reasons why people who distrusted Google or Dropbox were still using them were money (free tier) and inertia/resource implications of making a change from default choices made early in life of organisations and, as one person put it, "Google Drive tends to win out over everything just because its just a lot more developed and easier to use for the non-technical individuals within the organisation."

The extent to which open source solutions such as Nextcloud could replace Google or Dropbox for file sharing and cloud replication depended to some extent on whether there was someone to champion the transition and help people get set-up; whether it was available for free somehow (e.g. via an installation managed by someone else) and what tools clients/contributors/partners were already using.

There were a number recurring issues that interviewees found problematic in working with mainstream cloud file-sharing options:

- File conflicts were poorly managed and caused problems, especially where multiple people worked on a file and Internet access/synchronisation was uneven between users or patchy in general-- "it would be really nice if the p2p protocols could share ownership between several different individuals in a way, so that we can collaboratively write a p2p document."
- Needing to be online to get work done in situations with unreliable Internet.
- Poor access to account security features such as two-factor authentication for collectively managed accounts (e.g. role-based set-ups) because they might be tied to one person's phone or email account.
- Encrypting on top of cloud storage to protect sensitive data was too difficult to implement smoothly.
- Insufficient granularity in setting up sharing permissions for the complex range of stakeholders these coops often engage with both internally and externally. These can range from clear-cut situations with small, stable core groups working with a large number of external customers to much fuzzier arrangements of co-op members with short and long term collaborators and partners. There are greatly varying needs for flexible and retractable access and versatile permissions schema that can be customised to work flows via e.g. projects or roles.

- People would also like to be able to transfer data and share data within their organisation without going through corporate cloud services. If this is achieved by substituting a protocol working only on the local network, care should be taken to replace the default off-site backup function that mainstream cloud services perform.

Open source options for file sharing such, as Nextcloud or Syncthing, do address many of these points in varying ways, but there is a usability barrier to set-up and subsequent familiarity with flexible configuration options. CoBox would face the same challenges.

Some groups had detailed procedures for setting up new members with keys and managing permissions in a very structured way via a shared server space, but this option is only practical for technically oriented groups.

Whilst the necessity for blind replication in inter-organisation mutual back-up was clear, the interviews also revealed a set of more flexible permissions requirements for file sharing internally, as well as a need for secure options for high trust contexts, often with role-based sharing and security being a common requirement.

Password Management

Less technically confident users often expressed a desire to improve their password management. The extent to which this topic was considered 'handled' varied enormously. Whilst Keepass is an example of an open source password manager with a secure approach that does not lock users to a specific software, the use of *any* password manager should be considered a win, with less favourable browser-based online password services sometimes offering the benefit of actual user-uptake. Ease-of-use in key management for CoBox might be a make-or-break consideration in delivering secure mutual back-up that is not vulnerable to catastrophic loss if passwords are lost or changed by individuals.

Tech Support and Sysadmin

Even in cases where people have the tech skills to maintain their own open source infrastructure or figure out their own systems administration, there may be a reluctance to do so as this is very much seen as extra work that does not bring income or is not accounted for in core activities. This was where people were most likely to spend money on outsourcing maintenance and troubleshooting tasks or to pay for third party infrastructure.

When it comes to supporting new tools there is an understandable reticence about exchanging things that are not functionally broken, even if they compromise data sovereignty and control, for things that might break or require a big adjustment.

The design challenges here are as much about facilitating people, as implementing software and addressing infrastructure. In all cases where an open source alternative to corporate cloud tools was implemented, this solution had a champion and a maintainer in or near to the organisation and involved someone helping people set it up e.g. mobile Nextcloud clients for calendaring etc.

Other Software Issues

Areas where people felt a notable lack of practical open source solutions suited to small scale operations and cooperative ways of working included time tracking, especially for project-based work, role-based / multi-user access in 2FA and password management, invoicing and accounts, and integrated calendaring. In some cases interviewees were simply unaware of all the options available, while in others extensive research had not turned up appropriate options.

Software solutions designed for larger/different enterprises are hard to adapt. The answer is not just to scale down things like corporate ERP (Enterprise Resource Planning) tools. Small organisations face IT challenges that corporate actors solve at scale and those solutions are a) usually not economical for tiny organisations; b) contingent on resource-hungry customisation processes (e.g. formal task analysis); and/or c) don't reflect the way they need to work. Rather than an all-in-one solution, small organisations preferring open source tools need to be able to mix and match existing tools with good support communities that provide access to data in standard and easily portable formats. Finding better ways for existing tools to interoperate would in many cases be preferable to building new tools from scratch as people complained about new tools popping up but not sticking around, making users reluctant to invest effort in them.

In some cases there was a tension between wanting flexible, standards-based open source solutions and wanting comprehensive integrated solutions that provided easy overviews and avoided double handling of e.g. calendar data. Overall people preferred to avoid closed ecosystems (e.g. Google or Notion) controlled externally, desiring the convenience without the lock in. Easier ways to navigate multiple options to mix and match standards-based tools that interoperate and serve the needs of each organisation individually might address this to some degree.

In coops with both tech and non-technical members there were quite often two distinct cultures regarding the work flows felt to be intuitive between technology-focused workers and and non-technical professionals. Despite some projects attempting to bridge this divide, these are likely to remain diverged for the foreseeable future, given that most software is built by people connected with technical workflows. Good support for non-technical professional tools that do not force an all-in approach with locked-in data is therefore desirable.

Notable observations about the current ecosystem

Many problems already have solutions

The number of stated needs or tech challenges that actually lack solutions gets a lot smaller once we exclude problems faced by one group which other groups had solved in some way or for which there are in fact existing tools. As a project that may intimately connect organisations via technology in ways that depend on human networks, CoBox has an opportunity to broker a sharing of practices either as a planned dimension of support and roll-out or as an independent resource developed from this research.

Promoting a diverse infrastructure

Whilst most of the coops interviewed relied on corporate infrastructure in some way for some things, several had more independent infrastructure for some large or small proportion of their activities depending on in-house expertise and the nature of the data in question.

The number of providers of independent tech infrastructure to this sector was, surprisingly, less than expected, with a number of organisations clustering around providers with reputations for values-aligned practices, or people-focused service. The value of personal connections and word of mouth likely plays a role in this. Unfortunately, this also implies that if one of these key providers were to suddenly shut-up shop the impact on the sector could be substantial, at least in the short term. Rather than competing with these providers, CoBox could ally itself with them in ways that improve the overall resilience of the sector and strengthens the position of some key providers. This could be achieved by, for example partnering with existing providers for roll out and ongoing support functions where they already have niche expertise which the CoBox team may not currently be set up to offer.

Other potential threats to this ecosystem stem from widely used corporate cloud products - such as Google Drive and Dropbox - unilaterally changing their terms, features or access. Examples of this include Dropbox's decision to remove support for Ubuntu users who encrypt their hard drives. If Google decided to withdraw or substantially restrict its current free tier of services this would hit a lot of organisations very hard. Whilst this does not currently seem a likely scenario, it is difficult to predict how these large businesses will respond long-term to changes in their own environment, for example recent moves towards enhanced regulation in some jurisdictions.

There are already complex networks of data-sharing

The current descriptions of inter-dependencies and use of third party services suggests that organisations already distribute responsibility for the data they deal with. Interviews with co-ops that manifest a tech-based relationship, such as provider or recipient of tech services or data handling, made apparent that these relationships were often seen as simpler than they in fact are.

Customer X may know that they store data with tech service provider Y but may be unaware that provider Y also interacts with infrastructure provisioned by A, B and C, or out-sources data to yet other services, etc. Working with prospective CoBox users to understand the existing complexity and opacity of the trust models they currently employ might make the prospect of mutual back-up a more intuitively comfortable prospect. Instead of their data going 'away' to 'somewhere', which is presumed to be safe on the basis of various forms of existing trust based on things like personal connections, reputation or contractual arrangements, CoBox could help make these inter-relationships more visible and lower the threshold for having or participating in one's 'own' infrastructure as well as contributing to sector-wide resilience. Mapping out these relationships, both as recipients and deliverers of such services - and how these networks of trust are constituted - could be a useful step in the design process to help clarify how a mutual back-up system like CoBox fits into this existing matrix.