



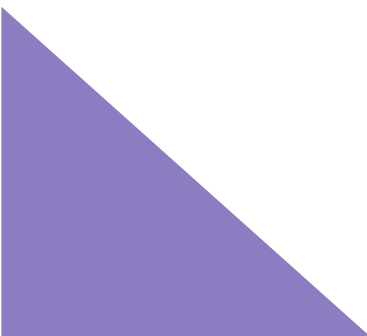
AiL

Clodagh McCarthy Luddy
Rosa Devine



Chosen brief: CSO

*A fundamental component of the CSO's business model is the **trustworthy** analysis of multi-industry and citizen microdata which has been shared by other public and private sector organisations. Mechanisms to verify the data has not been altered or tampered with during transmission could be improved.*



Problem in context:

CSO receives, handles and transmits large amounts of data. Trust in the integrity of the data they publish is paramount. While there are a number of allowable, purposeful instances of individuals at the CSO amending received data (anonymising, organising etc) unintended, unsolicited or unknown processing is undesirable and could, if it happened, seriously undermine the organisation.

Core:

Can you ensure data which may be changed over time has only been changed in allowable ways and for legitimate purpose.

Shopping list

Croissants
Ice cream
Cupcakes

Claire sent me her shopping list!

I'll buy those.

Hmm.
Lemme check
in with Claire.

Shopping list

~~Apples~~

~~Granola~~

~~Lettuce~~

Croissants

~~Eggs~~

~~Orange juice~~

Ice cream

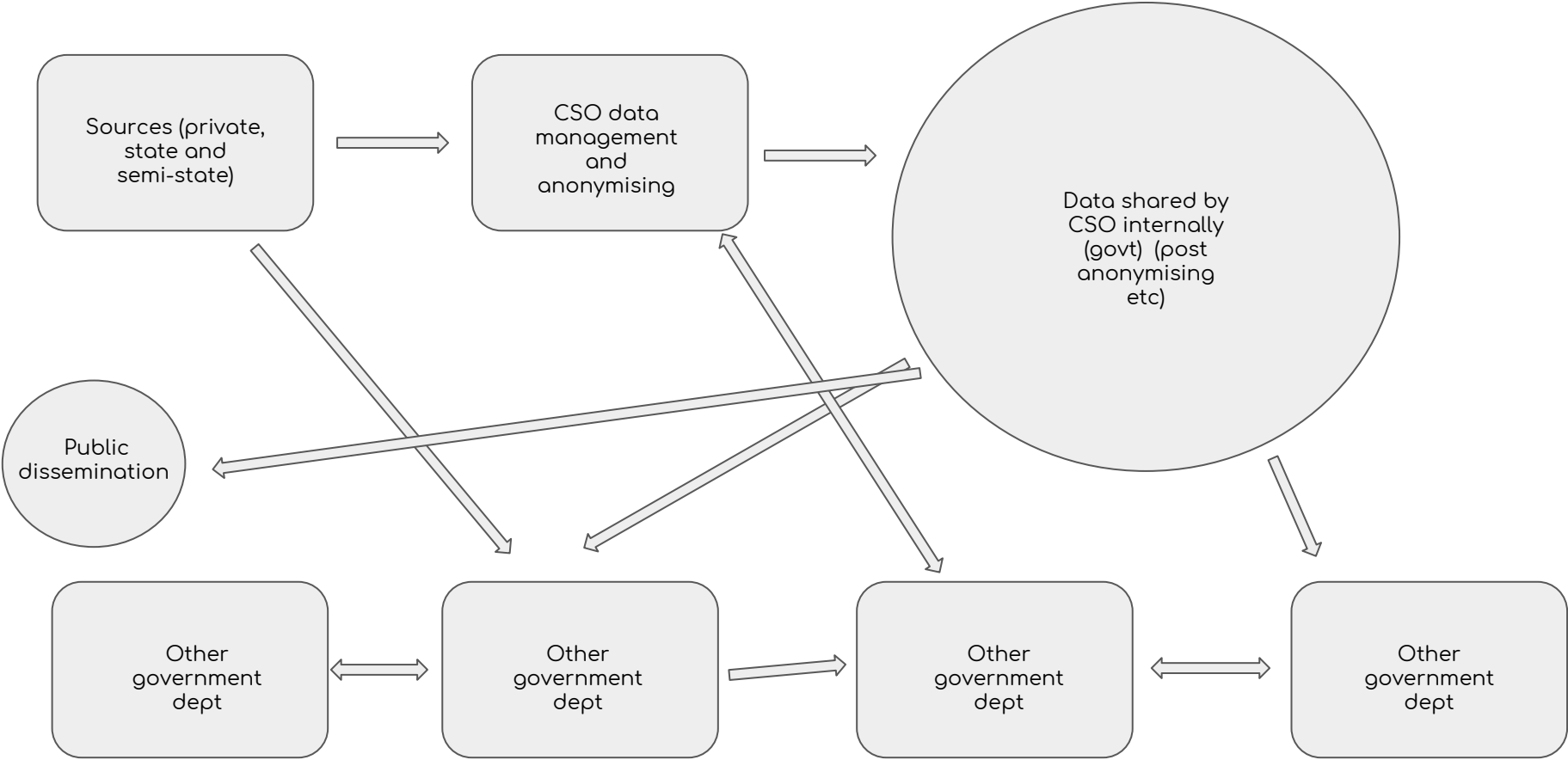
Cupcakes

File created:
000_responsibleAdult

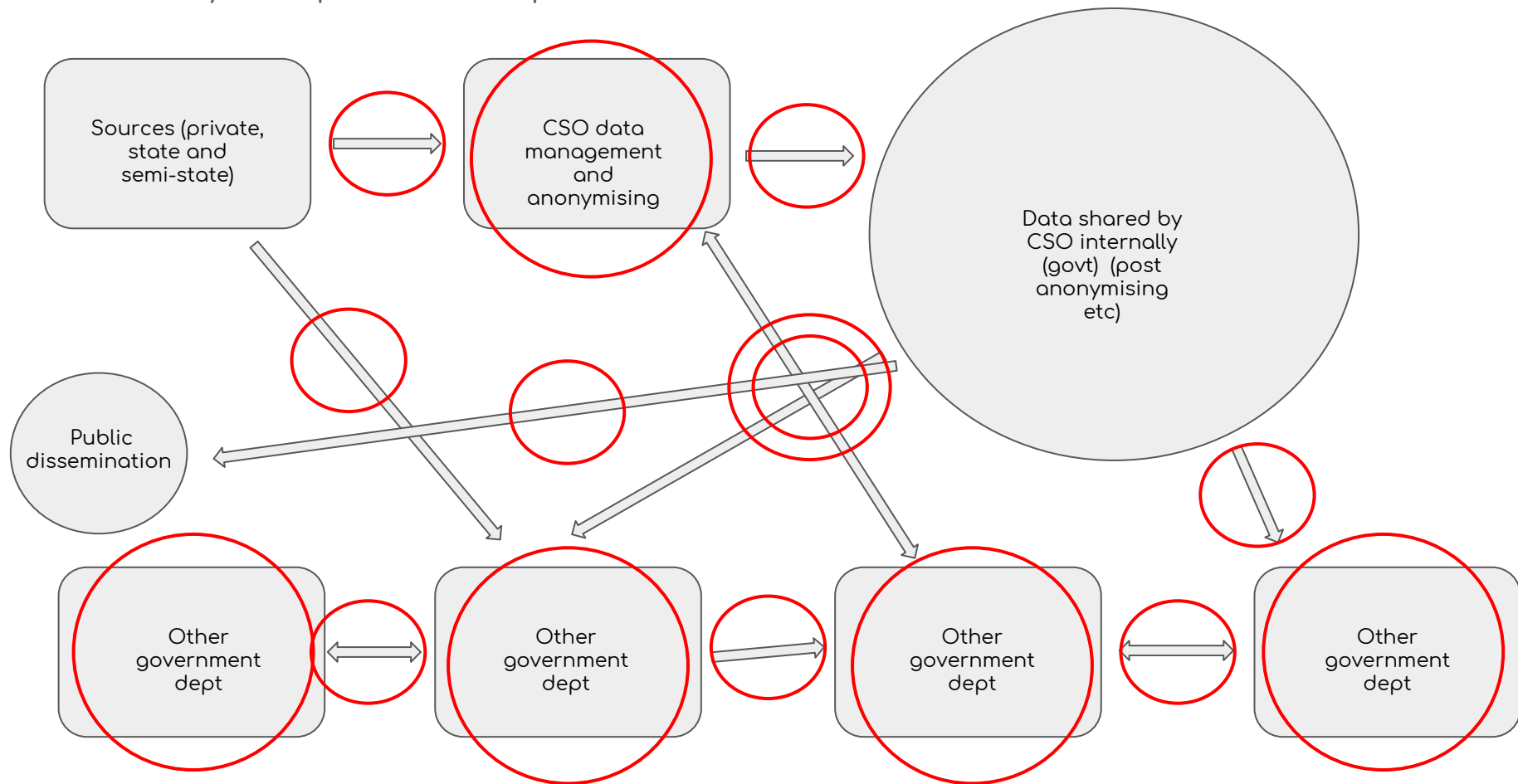
FileEdits[approved]:
1149_omlettemaker
832_responsibleAdult

FileEdits[insufficient]:
01938_impulseShop

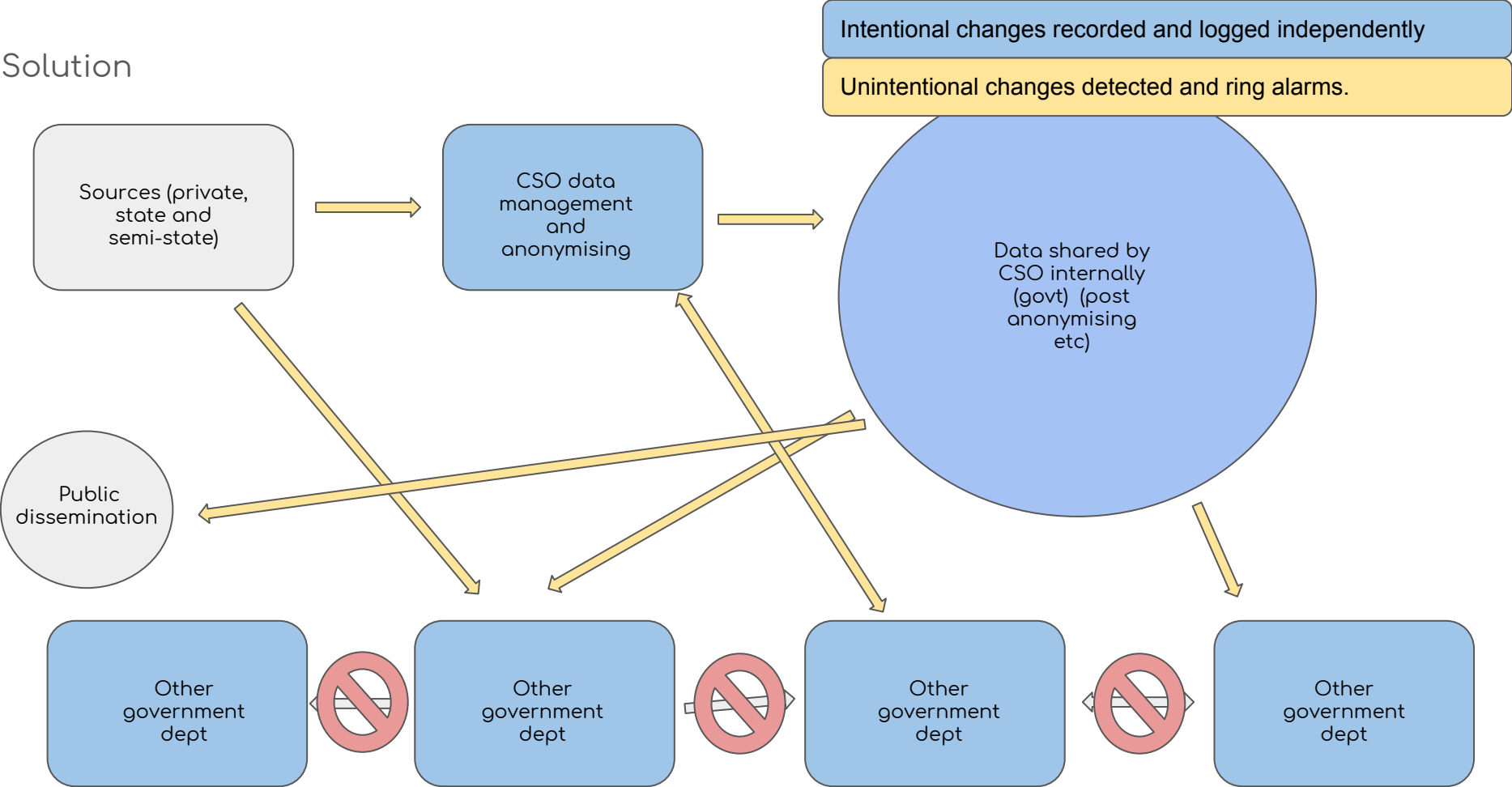
Current processes



A brief analysis of potential fail points



Solution



Solution

We propose a system that would use an independent programmatic third party notary to establish a data supply chain of all data handled by the CSO. This would:

1. Record changes to data in a transparent, independently verifiable way.
2. Allows data users (employees, researchers, business) to understand very specifically how the data has been processed on its journey to them.
3. Facilitate the speedy identification of non-standard intervention in the data which may merit investigation.

CORE CONCEPT:

There can be independent verification of data supply chain authenticity, without any access being granted to sensitive data and without impeding legitimate manipulation of data.

What do we know about the tech in place?

- FTP file transfer for inputs
- MS SQL Server shared storage
- Data analysis tools for interaction - R, SAS
- Ownership of data set latest state is an internal challenge - data silos
- Employees are trusted and educated with best practices for data collection, aggregation and management: GSBPM

Does blockchain apply?

1. Do we have a use case that doesn't fully fit inside one organisation?
2. Can this already be solved with existing tech?
 - a. Pure implementation
 - b. Cost and performance

Well,

1. We have many data stakeholders (citizens!) outside of the organisation that holds that data.
2. Partially, but we can improve on that
 - a. The problem statement is that the internal systems work, but they want improvements
 - b. Our addition is cheap, compatible with blockchain consistency speeds, and has been in production for 6+ years

Solution design

Use any blockchain as a third party notary - prove data integrity at a given time, without exposing it.

Leverage the internal systems in place to detect file changes, and sign the new state of the file publicly at each point.

Without access to the MS SQL system we have mocked another datastore and will demo pulling and manipulating files from that.

We have chosen for our demo to use a mature third party option and the Bitcoin blockchain for notarisation. This code is open source and the chain in use is not a requirement.



(Powering third party notary services since 2012.)

A note on the GDPR concerns raised...

These commitments to a third party notary are created using lossy hash functions. They contain no sensitive data at all, while still adding value.

Quick Demo

1. Portal with public listing of timestamps
2. Upload a file to the Administrative Data Centre
3. Play a service bot for a day

Try it for yourself [here!](#)

And find the scripts, tools and instructions for the service [here!](#)

Some considerations of this approach

1. Garbage in, garbage out. Bad input data will still be bad input data.
2. We're working with blocktime, not realtime.
3. It's not a panacea, data can still be modified. There are simply now more eyes watching both the current state and the historical record.

The business case:

Integrates with existing data management practices within the CSO, extends progress towards best practice and centralised file sharing aspirations.

Independently verifiable (CSO is itself independent for this reason), thus bolstering both transparency and trust.

Uses straightforward, well-proven and cheap technology.

Very low requirements for staff training or other barriers to implementation.

Maximum utility for minimum investment.