4.2 Use case 2: Registry

Registries are authoritative collections of information, often managed by government agencies. A registry holds information about a class of entities. Examples of such entities include individuals, businesses, species and organisations. In Australia, familiar registries include the immunisation registry, the business name registry, and land title registries. There are also well-known international registries such as the Domain Name Service (DNS). Some government registries are described as 'public', and can be queried by individuals. However, query access to these registries may be limited to prevent attempts at re-publishing or mining the data. Unfettered data mining could threaten commercial or personal privacy, and is often restricted using regulatory policies, and technically with query rate limiters and user access controls.

Some government registries contain periodically published, open data. In Australia, these are published through data.gov.au. In this case study, we specifically consider the use of blockchains for managing an open data registry of data sets, data sources, and data analytics services. So, we do not consider confidentiality or privacy issues for this use case. Blockchains provide transparency about their entire transaction history to all processing nodes. In a public blockchain, this means that the information is openly published. It is possible to run a private blockchain hidden behind a web service or other interfaces. This could limit access to the registry in a way that satisfies an appropriate access policy. However, many of the benefits of using a blockchain would be foregone in such an architecture. Private blockchains may provide a way to integrate registries across multiple government agencies, but this is not explored further below.

Although here we discuss open government data, we note that there are also non-government open data sets of national importance. This can include scientific data from universities, and data from non-profit institutions (including industry associations and consumer organisations). These data sets are not included in sites such as data.gov.au, but a blockchain of open data could provide neutral ground to federate references to all of these data sets. Also, instead of storing the open data directly in the blockchain, only metadata is stored. This provides a federated index to the data which are kept in the source repositories independently-managed by their governing bodies.

4.2.1 STAKEHOLDERS

For open data, the major stakeholders are data providers, data consumers, and the data registry. Data providers may include government agencies, research institutes, universities, and companies. Data providers record metadata about their datasets on the data registry, and make their data available on their websites. Data consumers query to discover datasets in the data registry based on the metadata. They can then download the datasets from the data providers for analysis.

4.2.2 KEY NON-FUNCTIONAL REQUIREMENTS

Some of the key Non-Functional Requirements (NFRs) for open data registries include:

- Integrity: each data provider should only be able to create and change registry entries for their own data sets.
- Availability: there should be high likelihood of being able to access the registry when desired, for both data providers and data consumers. This particularly applies to national public registries, which form the basis for many other services that utilise the data from the registries.
- Read latency: data consumers may need to repeatedly query the registry while browsing and searching for relevant data sets. This may be done programmatically from a graphical user interface and so should have low latency.
- Interoperability: A registry may reference other registries to reduce duplication and errors.
- Ease of integrating new data providers: to grow the network effects of the registry as a data portal, it is important to have low barriers (time, cost, and administrative burden) to add new data providers to the registry.

4.2.3 DESIGN OPTIONS

We provide three illustrative design options for such a registry. These are: conventional technologies operated by a single agency, a shared private blockchain operated by data providers, and a public blockchain.

Design 1: Conventional technology

Data portals such as data.gov.au implement a dataset registry using conventional technologies such as CKAN[5]. The CKAN software is run and managed by a single government agency. Data consumers interact with the registry to discover datasets, but retrieve datasets directly from data providers. The data providers may perform some permission management for data access independently. An illustrative high-level design is shown in Figure 6.

In the ecosystem of CKAN, the datasets in different CKAN repositories refer to each other through importing metadata from the referred repository to the primary repository and transferring it to the format used by the primary repository with possible customer-defined fields.

Design 2: Data registry on consortium blockchain across data providers

One design alternative using blockchain is to replace the backend of a conventional registry implementation with a consortium blockchain across data providers. Not all data is stored on the blockchain. For example, the registry may maintain a separate database for administrative purposes for permission management. As above, the data providers perform some permission management for data access independently. Instead of integrating with the registry's web service as in the conventional approach, data providers must instead integrate with the shared blockchain. Data consumers access the registry through an open data portal, which is hosted by a government agency. An illustrative high level design is shown in Figure 7.

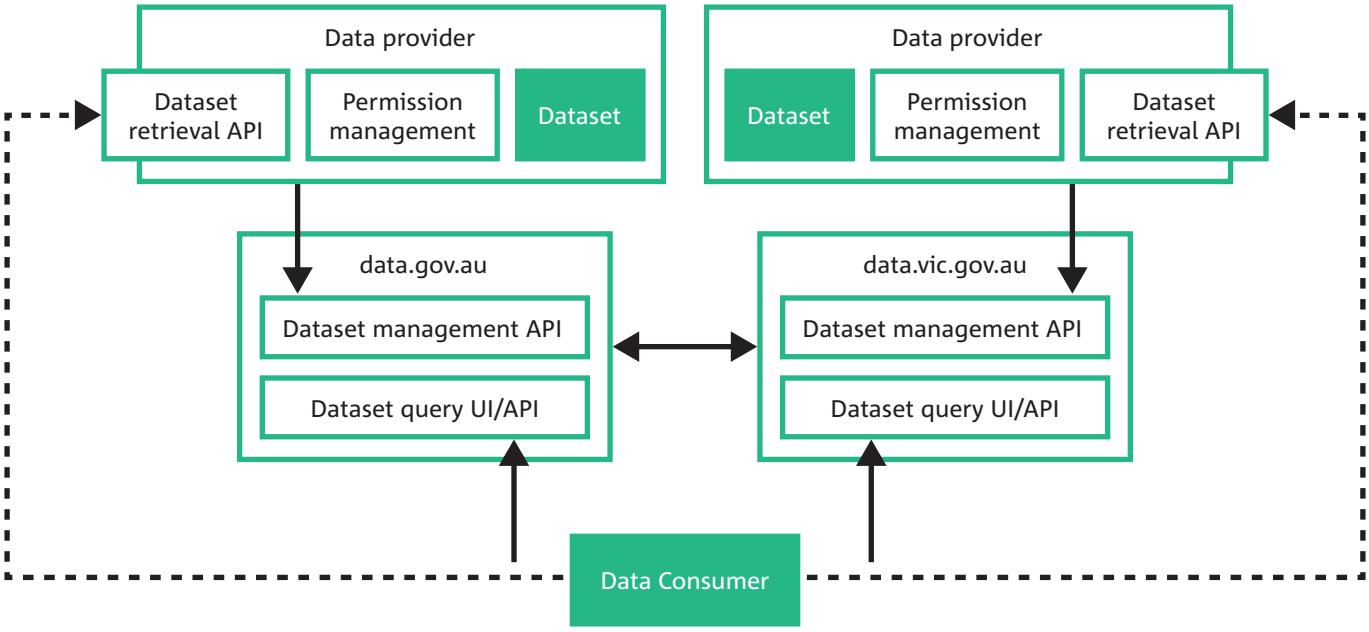

Figure 6 Design for a registry using conventional technologies, operated by a single agency

5 ckan.org

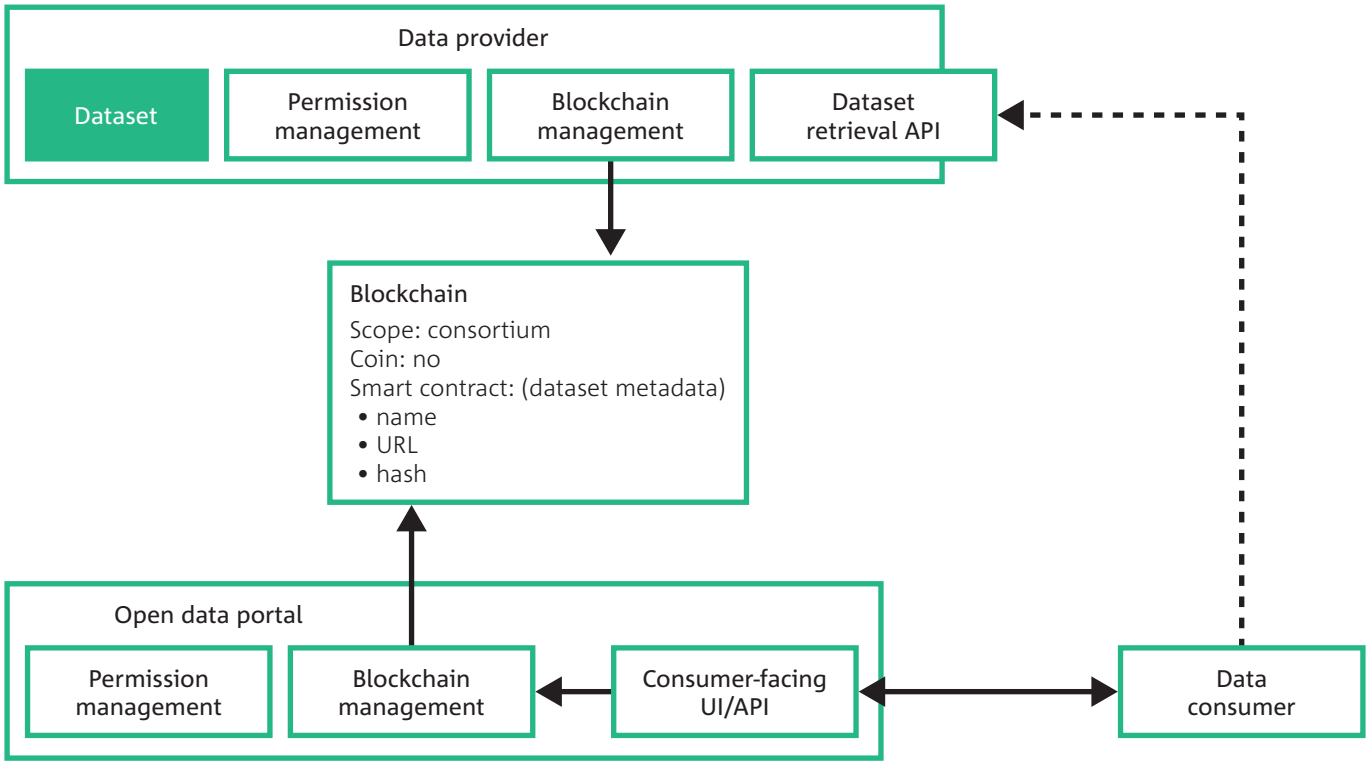

Figure 7 Design for a registry using a private blockchain

Design 3: Data registry on public blockchain

Finally, we consider a design which replaces the registry with a public blockchain. In this design there is no agency operating the registry. Instead the data providers independently record metadata on the public blockchain and perform their own permissions management and access control for their data sets independently. Note that there may still be an agency leading governance for the registry. In this design, data consumers are required to interact directly with the blockchain, rather than with a consumer-facing user interface or API. However, those consumer interfaces may be provided by a variety of commercial or personal systems, depending on the data provider's preferences. An illustrative high level design is shown in Figure 8.

4.2.4 NON-FUNCTIONAL PROPERTIES

Integrity

Design 1 relies on the registrar to create registry entries on behalf of data providers. New registry entries are validated solely by the registrar. In designs 2 and 3, registry entries can only be created by the data provider, using their private key, which must be kept secret for this purpose. All transactions are validated by all processing nodes in the blockchain network. In design 2, data consumers only access the registry via an interface which could modify information reported to consumers. In contrast, in design 3, data consumers hold a local copy of the blockchain, through which they access the registry, which removes the interface from design 2 as a possible point of manipulation.

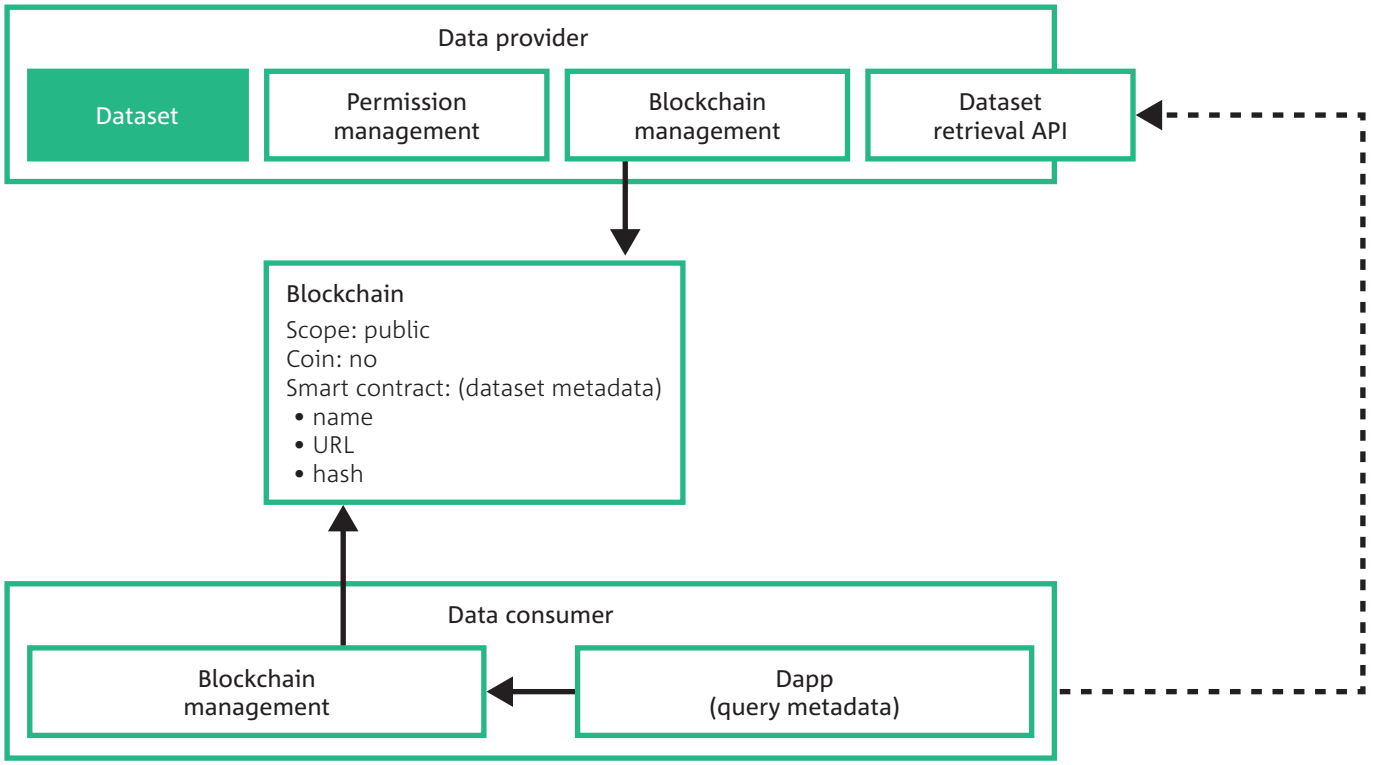

Figure 8 Design for a registry using a public blockchain

Availability

In design 1, the data registry system is a single point of failure for registry availability for all stakeholders. In design 2, the open data portal is a single point of failure for data consumers. Such a single point of failure could be mitigated with an IT architecture using redundant servers and network infrastructure. The use of a blockchain allows increased data redundancy which can improve read availability for data consumers. In this use case, write latency is not critical, and makes it easier to achieve higher service availability for writing registry entries.

Interoperability

In design 1, the datasets in different CKAN repositories refer to each other through importing the metadata from the referred repository to the primary repository and transferring it to the format used by the primary repository with possible customer-defined fields. Designs 2 and 3 use a blockchain as shared infrastructure, which means different registries can more easily interact with each other.

Read latency

Reading in design 1 and 2 is performed through a remote API over the internet. Compared with design 3, this is slower: in design 3 a blockchain local node is collocated with the consumer's query interface, and reading is done locally.

Ease of adding providers

In designs 1 and 2, new data providers are added through account creation and network configuration for the registry back-end services. In design 3, new providers can join by independently creating a new public/private key pair. Authentication of their public key could be certified by the registrar on the blockchain, or separately off-chain. Data providers in designs 2 and 3 must integrate with the blockchain, and should ideally run a blockchain node.

Cost of using public blockchain

To investigate the cost of using a public blockchain for this use case, we built a CKAN-inspired registry on a laboratory deployment of the Ethereum blockchain. The registry was populated with data taken from data.gov.au[6]. The example registry has three entities: organisation, package and resource. Each entity has 6 attributes that are stored on the registry. Architectural decisions can affect the cost of deploying and executing the registry. For example we could use a 'single' registry that holds all records as values in the data store as a singleton smart contract, or we could use a 'distributed' registry which manages each record as a separate smart contract. For a 'distributed' registry, a main registry smart contract creates entry contracts and stores pointers to them. The 'single' option is suitable for simple registries, while the 'distributed' option is suitable for registries with complex operations, such as finer-grained permission management at the level of individual records. Table 1 gives statistics about the different options, both for creating a registry on the blockchain and the cost of adding a record to the registry.

The cost of creating a registry contract is comprised of fixed costs and variable costs. Fixed costs include the base amount for the transaction itself and the cost for allocating an address on the blockchain. Variable costs are affected by the architectural design of the registry contract, for example, the cost of data payload. Similarly, the cost of adding records to a registry is also comprised by a fixed cost for the transaction itself, and some variable costs including for the data payload and cost to execute the functions defined in the registry contract. Compared with conventional databases, using public blockchain costs more to add records. However, the data becomes globally replicated and the blockchain ecosystem will retain this data indefinitely as long as the blockchain exists, at no additional cost.

Table 1 Gas cost and dollar cost of registry functions [7]

	NUM.	REGISTRY DEPLOYMENT				RECORD CREATION (AVERAGE)			
		GAS COST		USD COST[7]		GAS COST		USD COST	
		SINGLE	DISTRIBUTED	SINGLE	DISTRIBUTED	SINGLE	DISTRIBUTED	SINGLE	DISTRIBUTED
Organisation	533	1836926	2542604	US$0.9	US$1.3	183266	931179	US$0.1	US$0.5
Package	33810	1836926	2542540	US$0.9	US$1.3	340022	1090174	US$0.2	US$0.5
Resource	64147	1777127	2548455	US$0.9	US$1.3	302041	1065760	US$0.2	US$0.5

6 Scraped at: 2017-03-07T15:59:32+1000

7 https://poloniex.com/exchange#usdt_eth