THOUGHT LEADERSHIP

Data Sharing

Matching and Routing

Delivering Transformation. Together.

sopra steria

Introduction

There are many reasons why data needs to be shared in the public sector, and in reality there are only a limited number of ways to implement a sharing solution. This paper covers the main data sharing solutions in a generic fashion. This view is based on a study undertaken of all the main NHS and social care sharing solutions in Scotland, and more recent consultancy work on data sharing between schools, police and health agencies.

The requirements

Any data sharing solution is driven by what data needs to be shared, and how. Some of the key requirements that will determine the solution are:

1. What is the distribution mode?, e.g. one-to-many, many-to-many, many-to-one, or one-to-one. Is the communication with a specified professional, a role, a team, a system or an agency?
2. Will data be pushed or pulled?, e.g. a push to registered subscribers, or will accepted organisations pull, request or read data on demand?
3. Are there any existing sharing solutions that need to be integrated?
4. Is there a requirement for receipts on successful or failed delivery, or when read?
5. What is being shared?, e.g. structured data and/or documents.
6. What are the expected data volumes and profiles?
7. What is the urgency?, e.g. high priority child protection alerts or lower priority shared plan contributions.
8. What level of data security protection is needed?
9. What personal identifiers can be used for matching, where are they held, and is their use restricted in any way?
10. Does there have to be a definite match using matching Ids, or is there a business tolerance for some level of fuzzy matching using demographics?
11. Is matching fully automated, or is there provision for manual matching of exceptions?
12. What are the business processes for a failed match?, e.g. return to sender or send to receiver.
13. What's the transport mechanism? e.g. a secure VPN or web based.

There are other considerations, but this is enough for starters.

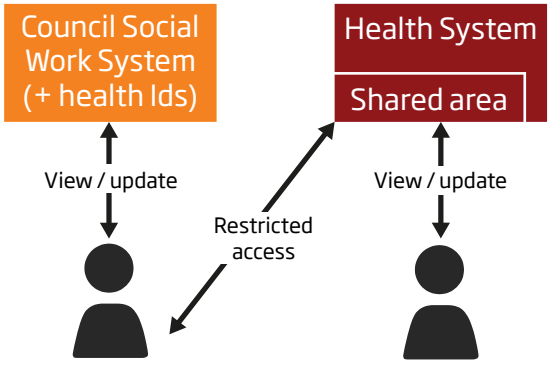
The solutions

Assuming the "what" is reasonably clear, then the "how" can be considered. Although there's a danger of over simplification, it's helpful to consider five different models for data sharing solutions. It should be noted that the features of the models can overlap, and one model can evolve into another, e.g. a stand-alone central store could evolve into an integrated central stored if the solution allows the development of electronic interfaces.

1. Single shared system
2. Stand-alone central store
3. Integrated central store
4. Data portal
5. Central messaging hub

Single shared system

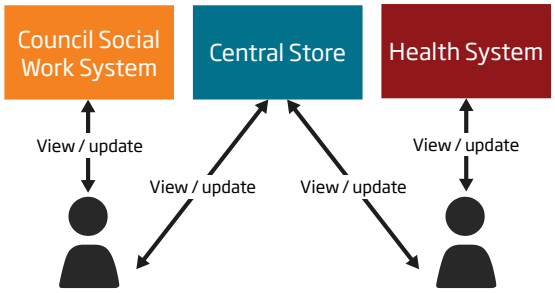
- Information remains in a single agency system with no links to other systems.
- Access is extended to other agencies, as long as security is sufficient, and is restricted to a limited subset of data with restricted search facilities.
- In order to assist with manual record identification linked systems may be seeded with a common individual Id, e.g. an NHS Id.



Single shared system

Stand-alone central store

- Centrally or cloud hosted database.
- Users can upload documents and jointly update data.
- No electronic links to other systems.
- Secure direct access from any location, e.g. web based.
- Separate logons from other agency systems.



Stand-alone central store

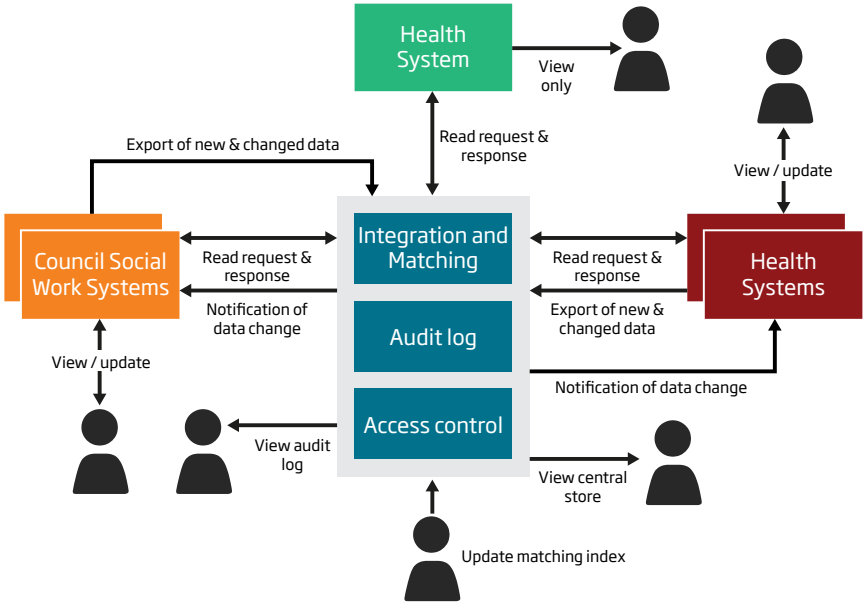
General tip

Any data sharing solution needs to be supported by agreed and documented data governance arrangements between the sharing partners, for example an Information Sharing Protocol (ISP). The solution and the governance are dependent on each other.

All sharing organisations also need to assess the risks to the data being shared (both at rest and in flight) and how they will manage these risks, e.g. prevention, mitigation or acceptance. The assessment can take the form of a Privacy Impact Assessment using a standard assessment scheme, for example the Government Security Classification scheme.

Integrated central store

- Uses a publisher/subscriber concept with a central data store and a central matching index.
- New, updated or deleted data is automatically or manually exported from the linked systems to the store. Notifications of changed data can be sent to the linked systems.
- A central matching index holds linked identifiers from the agency systems. Matching is automatic with possible manual backup. The index is manually and automatically maintained.
- Access can be limited to the service user context in the linked agency systems (which gives greater access control), but direct access to the central store is also possible.



Integrated central store

General tip

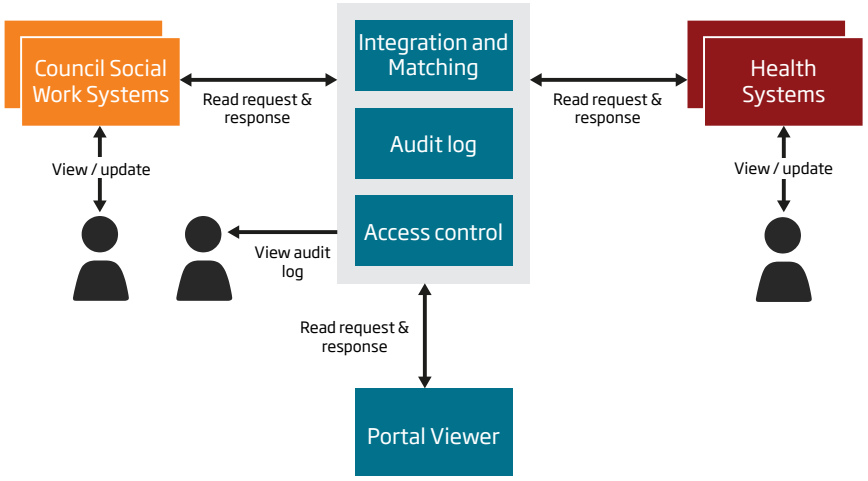
The implementation of data sharing should be an ongoing incremental process, with new types of data being shared between different organisations. As such it is useful to be able to track the status of data sharing between organisations, as shown below.

Types of data shared	Organisation		
	1	2	3
Child protection messages / alerts	L		L
Assessments and plans		B	B
Integrated chronologies	D	P	P
Associated professional details	D	L	L

L	Live solution
B	Solution in build
P	Planned solution, i.e. firm plans or budget assigned
D	Desired solution – wish list stage
	No known requirements

Data portal

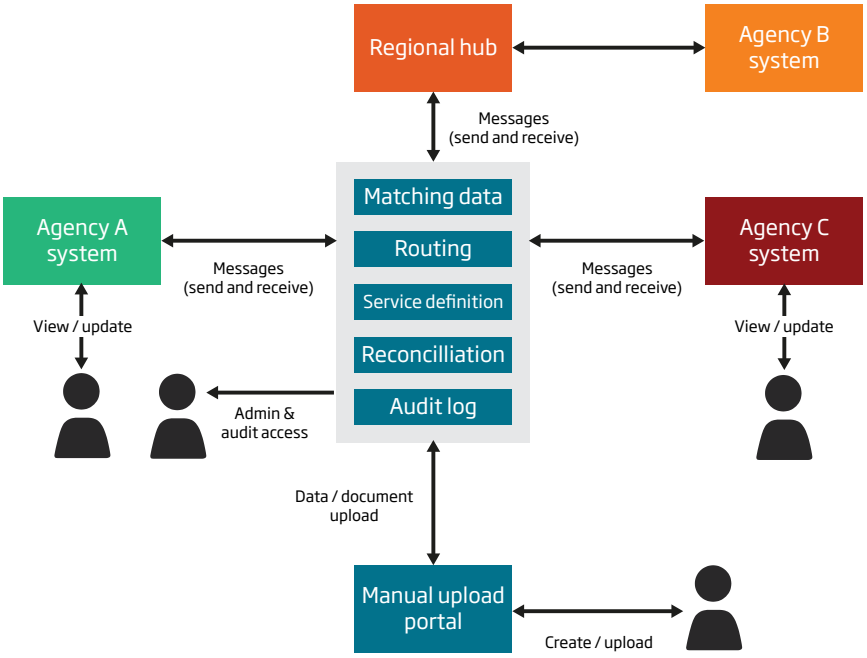
- A portal uses a request / response model to provide a unified view of data relating to an individual from several linked systems.
- There is no central data store or matching index - all the linked systems have to be seeded with a common individual Id, e.g. an NHS Id. This may involve an initial automatic and manual seeding exercise, followed by ongoing seeding as new records are created.
- The data "owners" of the common Id need to agree that the Id can be held in an external system, and that acceptable data security and governance arrangements are in place.
- External system users need to be able to easily obtain the common Id, and processes are needed to monitor and enforce data quality.
- Access can be via the service user context in a linked agency system, or by logging directly onto the portal. Direct logon may provide a list of the user's cases, or a search page. The search could be by an individual Id or a minimum number of demographic details (in order to restrict "fishing" searches).
- Users have access to restricted sets of data depending on role.
- Direct access to the portal supports facilities such as viewing of audit data and access control.
- A data sharing portal can be implemented by extending an internal multi-system portal to integrate systems in external agencies.



Data portal

Central messaging hub

- The hub provides matching and routing functions only – apart from a matching index no personal data is held.
- A lightweight hub will hold minimum persistent data, alternatively a heavyweight hub could hold more persistent data, for example requested data could stay in the hub for a defined period of time in order to support repeat requests in a short time span.
- Both deterministic matching (on linked Ids in a master index) and probabilistic (fuzzy) matching can be provided. Deterministic matching assumes the presence of a single field that can give a reliable match between two records. Probabilistic matching involves the comparison of several items of demographic data between two sources (e.g. name, date of birth and gender). Every item is assigned a weight that shows the degree of matching. The total of the item weightings indicates the probability of a match. Sharing agencies have to decide what probability they are prepared to accept as a good match.
- Organisations interface via a small flexible API providing input and output services.
- Routing can be as complicated as needed, for example, one-to-one push or one-to-many request.
- Messages comprise encrypted routing and identity metadata, with a separately encrypted payload that is inaccessible to the hub.
- A secure web based portal can be provided to allow manual input of data or documents for matching and routing by the hub.



Central messaging hub

Common functions

All the models described show common functions (e.g. matching) and qualities (e.g. flexibility) that need to be considered and implemented by any data sharing solution:

- Record matching – Personal records from separate systems need to be quickly matched with a high degree of certainty. As far as possible this needs to be automatic with minimal manual intervention. This can be by:
  - A single common Id held in all the systems, e.g. an NHS number.
  - Different Ids from different systems matched and held in central store
  - Matching of good quality demographic data.
- Data routing and protocol support – A solution can simply route data based on information set by the sending organisation, or more complex routing and message protocol rules can be built into the solution.
- Auditing – Any solution needs a comprehensive audit log of data updates and views, with a means to quickly report on the log and identify misuse.
- Management reporting – As solutions mature they become a good source of management information, e.g. identifying differing levels of sharing and trends.
- Secure links – All solutions need a secure means to link systems. Often this is based on links between NHS and Local Authority networks. However further sharing with Police and other agencies may require additional links. Links need to have sufficient capacity, availability and reliability. Giving access to non-government actors such as the Third Sector, private companies and the public presents further challenges. A system-to-system architecture may not be suitable and alternative secure sharing solutions may be needed (for example the manual upload to a Hub already mentioned).
- Scalability - Solutions need to be scalable to allow the integration of many agency systems, with different matching and routing requirements.
- Flexibility - Solutions need to be flexible enough to support a variety of constantly changing business practices and legislative change. They also need to be able to support upgrades and replacements to the interfacing systems.

Conclusion

Implementing a data sharing solution can often be difficult, with complex technical and organisational challenges to overcome. However an understanding of the common requirements and generic solutions can make the task more achievable by helping organisations to fully understand their goals and to select an appropriate means to achieve them.

About Sopra Steria

Sopra Steria, European leader in digital transformation, provides one of the most comprehensive portfolios of end to end service offerings in the market: Consulting, Systems Integration, Software Development and Business Process Services. Sopra Steria is trusted by leading private and public organisations to deliver successful transformation programmes that address their most complex and critical business challenges. Combining high quality and performance services, added-value and innovation, Sopra Steria enables its clients to make the best use of information technology.

How we can help

As a systems integrator Sopra Steria can assist with all the main strands of a data sharing implementation, including:

- Requirements analysis
- Data governance definition
- The design, build and implementation of interfaces between packages and system
- Data security assessments and solutions
- End-to-end test planning and execution
- Master data management

We can also utilise our general experience of building databases, inter-system interfaces, and data extracts and loads. Our technical expertise is complemented by our in-depth business understanding of the sectors typically involved in data sharing, i.e. health (both primary and secondary), local government, justice and education.

SOPRA STERIA
Three Cherry Trees Lane, Hemel Hempstead, HP2 7AH
+44 (0)370 600 4466 - info.uk@soprasteria.com

PB038V01

www.soprasteria.co.uk