DDSF Project 1a
Standard Safeguarding Dataset
Final Report

Report Author: Alistair Herbert, Data to Insight

Contributors: Paul Dryden (Hertfordshire County Council), Robert Harrison (Data to Insight), Joseph Cornford-Hutchings (Data to Insight)

Report date: April 2024 – [Draft 1.0]

Contents

1. What is the Standard Safeguarding Dataset?

In 2022 the Department for Education (DfE), as part of its response to the MacAlister review of children's social care, launched a Digital and Data Solutions Fund (DDSF) to develop sector-led responses to significant challenges and opportunities in the children's social care data landscape.

DDSF1a, as specified by the DfE in consultation with the sector, is a project to introduce a new Standard Safeguarding Dataset which is broader in scope than existing statutory data returns, more useful to local authorities (LAs), and easy to deploy for LAs using any of the major current case management software solutions. The project is led by Hertfordshire County Council as part of a consortium of LAs in partnership with Data to Insight (D2I), the sector-owned service for local children's safeguarding data work.

The project presented the first "live" version of its dataset specification in financial year 2023-24, alongside methods for producing the dataset from the market-leading case management systems.

2. What is this document?

This document presents an overview of the project history, its outputs (including separate detailed reports and technical specifications), and its possible futures.

This document is intended as a single signpost to the several distinct project outputs to date, and as a reflection on the project's successes and future potential.

3. Executive Summary

In response to the 2022 independent review of children's social care ("the MacAlister review"), DfE initiated a whole-system response in the form of a consultation response, revised practice guidance, and a range of innovation projects across practice and enabling technologies. The Data and Digital Solutions Fund (DDSF) marked a significant element in this initial response, formalising government's commitment to "put Local Authorities in the driving seat of change".

The Standard Safeguarding Dataset (SSD) represents one of the DDSF's major products: a sector-designed dataset specification for production by any LA, which can improve sector collaboration, help LAs generate insights, and ease data interoperability between local and national government.

To deliver the project, we made careful use of the DfE's initial specification (see report section 4, below), and developed this further using two phases of extensive user research which we documented in separate reports (sections 6.4 and 6.5). The project was a major collaboration between LAs across England, and in this report we tell the story of that collaboration and the project milestones (section 5).

The final product comprises several separate code outputs and documents (section 6). The key resources are as follows:

- **Code and guidance for the specification is located at the Data to Insight code repository, hosted by GitHub: GitHub - data-to-insight/ssd-data-model: Standard Safeguarding Dataset (SSD).**
- **The live specification is hosted for user navigation on GitHub Pages: data-to-insight.github.io/ssd-data-model/.**
- **Code methods for producing SSD data extracts from CMS products are, at the time of writing, stored in a private repository within the data-to-insight (github.com) code library.**

- Key findings from initial user research are summarised in section 5.2, and detailed in **a separate document "Summarising User Research" (June 2023).**
- Key findings from secondary use research are summarised in section 5.2, and detailed in **the full report "SSD New Data Workshop Synthesis" (October 2023).**
- **A separate "Future Support Plan" (February 2024) details our proposals for maintaining the SSD in 2024-25.**

The SSD's successful delivery presents a range of opportunities for further work or interactions with ongoing initiatives in the sector, including relationships to other DDSF projects, possible initiatives to benefit LA data work, and benefits to DfE objectives (section 7).

This report concludes a year of intensive and productive work to introduce a first live implementation of the SSD. Later sections of this report detail our approach to assessing how well we met our original objectives, and what work would be most useful to pursue next (section 8), key learnings from the project for colleagues involved in other cross-sector data work (section 9), and acknowledgements of the many partners without whose unique contributions the work could not have arrived at its current stage (section 11).

We hope colleagues will find the SSD useful, and this report valuable as a record of the work.

4. The brief

The DfE provided a clear brief for this project which we repeat in full below, with our emphasis:

> LA data teams must balance their limited capacity with high demand for data work. One way they address this is by sharing data tools, tasks and methods across professional regional and national networks. However, such **economies of scale can only be achieved where datasets are standard across LAs**. This limits data collaboration between LAs to standard 'core' datasets – most notably the Children in Need (CIN) Census and SSDA903, for children in need and looked after children respectively – and Ofsted's "Annex A" specification of child-level data to be provided during an inspection. These datasets lack much of the depth and granularity that individual LAs will look to incorporate in bespoke local analysis, including to understand quality of practice.
>
> This leads to a second issue: while it is inevitable that LAs differ in the quality of their data work to some extent, the lack of data visibility across the sector leaves some LAs unaware of what they could do to improve data collection and analysis. **The existing standard datasets do not incorporate the breadth and depth of information which more data mature LAs analyse**. We believe that this knowledge gap inhibits improvements in analytical practice in the sector and so limiting better outcomes for vulnerable children.
>
> **We are therefore looking for a partner to develop a standard CSC dataset that is broader than the previously mentioned core data sets and solutions that would enable the production of this of this data set across all LAs.**
>
> **Stage 1**:
>
> - **mapping of data** that a broad range of LAs are using to understand their services **beyond those used 'core' data sets.**
> - **user research across a broad range of LAs to define the data items, data structures and production methods** which will best serve the sector drawing, where relevant, from existing research on this topic proposals for how other safeguarding partner information can be incorporated into the standard CSC dataset.
> - documentation explaining the findings of user research, rationale behind the chosen indicators and how they're intended to be used.
> - **a standard CSC dataset specification** (including a list of data items and definitions) incorporating multi-year data histories across the range of safeguarding activities performed by councils, significantly expanding on the available standard datasets (e.g. CIN Census and Ofsted's Annex A) in line with user research outputs.
> - identify where data items in the proposed data set are already included in the CIN Census and SSDS903, Ofsted Annex A and the Regional Improvement and Innovation Alliance quarterly data return.
>
> **Stage 2**
>
> - **a suite of standard easily adoptable methods for regularly producing the standard CSC dataset** from each of the case management systems currently used by safeguarding authorities, either as independently developed, free, open-source products, or as integrated components of those case management systems at no future cost to customers.

- **a framework for maintaining and developing the standard CSC dataset** into the future so it will be accepted by as many LAs as possible, and include changes made as a result of the care review and social work practice, keeping the interests of vulnerable children at the heart of all design decisions.

We wholeheartedly endorsed the DfE's vision for this project, and adopted the brief's separation of two stages, resulting in two key project delivery strands which are nevertheless closely related.
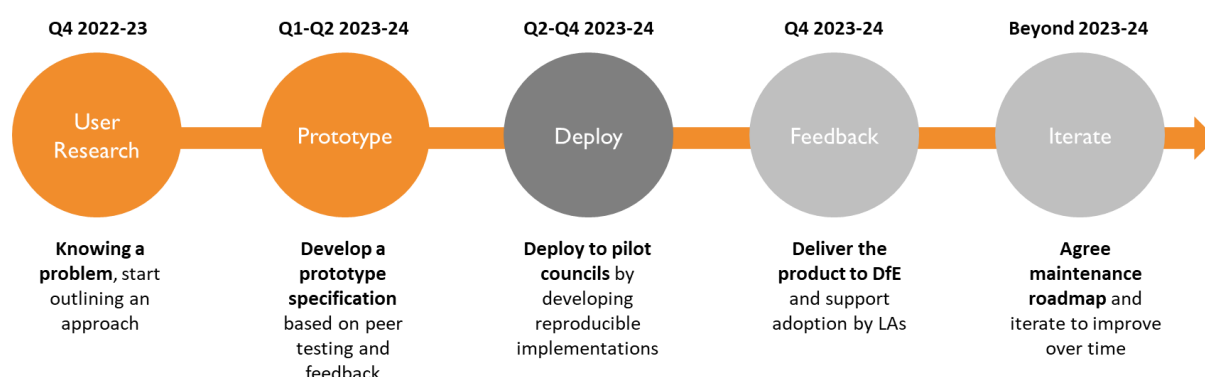
5. The project year as it happened

Our project commenced with a successful bid to DfE's Data and Digital Solutions Fund, sponsored by a partnership of LAs comprising:

- Hertfordshire County Council
- Essex County Council
- Knowsley Council
- North West ADCS
- East Sussex County Council (host LA for Data to Insight)

We also established a wider circle of contributing partners including Manchester City Council, Trafford Council, Sutton Council, Newham Council, and the London Innovation and Improvement Alliance.

5.1. Initiation

We developed a **high-level plan** for the year based on our successful bid, to which we fitted our detail work. The original high-level plan changed very little after the project started; with minor exceptions to be discussed below, we met our initial target dates throughout the project's year.

Q4 2022-23	Q1-Q2 2023-24	Q2-Q4 2023-24	Q4 2023-24	Beyond 2023-24
User Research	**Prototype**	**Deploy**	**Feedback**	**Iterate**
Knowing a problem, start outlining an approach	**Develop a prototype specification** based on peer testing and feedback	**Deploy to pilot councils** by developing reproducible implementations	**Deliver the product to DfE** and support adoption by LAs	**Agree maintenance roadmap** and iterate to improve over time

We identified and secured key project management resources and sector expertise within partner councils, and commenced **successful recruitment** of two technical posts to be managed by D2I, with the core team in place by 1st May 2023 and a robust project governance framework in place with

support from Essex County Council. During the recruitment phase we also produced our initial plans and outline documents.

We leveraged existing resources within D2I to commence our user research during our technical recruitment period in the final quarter of 2022-23, ensuring that our new technical team would have access to relevant sector insight as soon as they started work. Our **initial user research findings** were delivered to DfE colleagues in May 2023, which helped us to establish **good open working arrangements** early in the project.

5.2. Research

Initial research for the project took two key forms:

- **User research** with LA experts and other sector stakeholders.
- **Desktop research** and analysis of existing data structures and standard reporting arrangements in the sector.

We split our user research between in-depth one-to-one **interviews**, to capture specific views, and broader open **workshops** to foster group discussion and interaction. We also delivered a **survey** to gather more specific feedback on individual areas of data practice.

While the respondent map (pictured) shows strong representation from LAs within the North West and London partnerships, we also see representation from almost every region nationally. This breadth of experience and expertise – not only in terms of geographical location but also the different kinds of council and different modes of organization – was crucial to developing **a picture of data work in the sector which reflected the interests of the whole sector**, not only the most data mature or well-resourced councils. Other LAs not noted in this map also contributed to the survey exercise.

We also engaged with the academic research community and with different teams within DfE, seeking to understand the purposes to which local administrative data may be put once it leaves the originating LA.

We spoke to more than 50 colleagues across these sessions, as well as receiving approximately 50 survey responses. We summarised **key insights from initial research** under the following headings:

- Most LAs want better data *and* most LAs place high value on benchmarking capability.
- Many LAs mistrust DfE use of their data.
- We're right to think current data doesn't do a good job of exploring outcomes/"voice".
- Expanding or changing data arrangements in LAs is costly.
- LAs aren't all the same.

As well as providing varied and valuable insight into local perspectives on the value of standard data work, our initial research identified **design principles for our dataset specification**, including:

- the need for a "broader and flatter" core data output than existing standard returns.

- the need for a distinction between "current" and "new" data items, i.e. those which should be included in our dataset specification and extract method, and those which are not currently sufficiently common across LAs but should be captured and communicated in a separate report to DfE.

A separate document "Summarising User Research" (June 2023) is available and provides full detail on the above findings.

Later in the project, we undertook a second research phase focused on the above-noted distinction between "current" and "new" data items. We wanted to better understand the sector's views on potential extension of standardised data work into new areas of practice, and the potential for improvements to areas where existing standard data is available but limited in scope. We spoke to almost 100 colleagues representing 38 LAs, and we organised **findings from our second research phase** into the following key sections:

- Items feasible for inclusion in the SSD v1, despite not forming part of a current statutory return (typically additional context within existing standard data domains)
- Items for future consideration when developing national standards (typically data domains which are not currently well-served by any existing national standard, or in some cases not collected locally in all LAs)
- Other overarching themes (including understanding outcomes data, disaggregating national data, and linking to datasets typically maintained outside the core CMS)

We also provided a set of prioritised recommendations for further action within the sector, including some discussion of the existing breadth of data development across the sector, and the limitations of current national approaches to data development.

The full report "SSD New Data Workshop Synthesis" (October 2023) is available and provides full detail on the above findings.

Between these two user research phases, we also performed **extensive desktop research**, thoroughly mapping all existing statutory data returns and other known standard datasets in the sector, and using these to refine our understanding of the current data landscape for LAs, and the possible structures for our new dataset specification.

5.3. Prototype specification

We deployed our first draft specification early in the project's life, indeed before we had completed our map of existing data items. **We used the live specification as a primary reference point for continuous engagement with LA colleagues** in the partnership and in other LAs. While desktop research continued, we tested approaches to presenting the specification, and identified a production approach which used Python to automate the generation of YAML data files, and associated web-based and PDF visualisations, from a local Excel information store.

Once in place, we established a routine refresh of the local working document to the public-facing dataset visualisations, allowing us both to test and refine our production approach, and make our work visible to LAs colleagues as early as possible, thereby generating opportunities for feedback.

As our three research strands generated insight, we interpreted this into the draft specification as part of our iterative process. We also expanded the metadata associated with each aspect of the specification – adding new visualisations in response to user feedback, and developing the dataset specification itself to include more information about data definitions, sources, and connections.
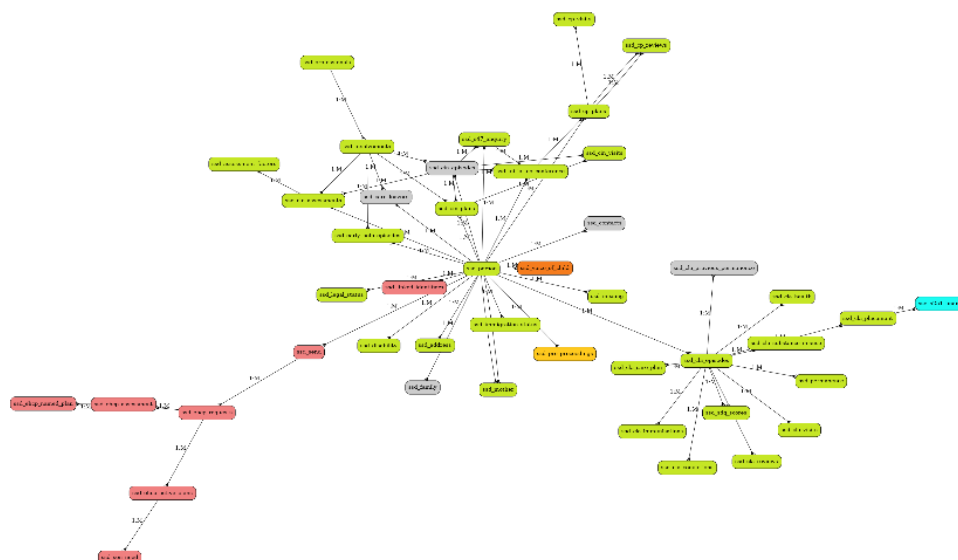
Code and guidance for the specification is located at the Data to Insight code repository, hosted by GitHub: GitHub - data-to-insight/ssd-data-model: Standard Safeguarding Dataset (SSD)

5.4. Pilot deployment

As research concluded and the draft specification stabilised (in autumn 2023), we **commenced development of production methods to generate the SSD dataset from local systems**. In our bid we identified two possible routes for this work: one was to work with CMS suppliers to build the SSD production code into their core systems and maintained business processes; the other was to work with LAs to produce generic SQL-based data extract scripts for each CMS which could be tailored to the local context in each user LA.

We identified that a key stipulation in the DSDF funding framework would inhibit the supplier engagement option; the funding agreement was explicit that all intellectual property generated by the project would be the property of the DfE on project close. After discussion within the steering group and with partners, we mapped the likely pathways for supplier engagement in light of these stipulations and decided that the risk of a compromised product, and the risk of initial negotiations delaying the overall project, was too great, and **we selected an LA-engagement option as our first choice of approach**.

We quickly found multiple suitable LA partners for co-developing SQL extracts from the two market-leading CMS products, and started work mapping our specification to their local systems. As this neared completion, we began to write and test SQL extract code locally, and then to test it against other partner LA systems to understand compatibility and possible tailoring needs.

Midway through this work we identified a database technology split for one of the market-leading CMS products, requiring us to work with another test partner and adapt our existing T-SQL code to a separate Oracle-SQL option for that product.

By March 2024, **we produced shareable data extract code for the first version of the SSD specification, for the two market leading CMS products**, including testing our pilot developments in unrelated partner councils to demonstrate national scalability.

Within the project year we were unable to complete the same work for three CMS products with smaller market share – we have outlined our proposals for completing this work in the separate "Future Support Plan" document.

5.5. Product delivery

Throughout the project **we embraced open working principles**, looking to share our work-in-progress widely, frequently, and in ways which suited our varied audience of stakeholders.

We blended formal show-and-tell meetings with open team meetings, including DfE in our regular project planning discussions to ensure client visibility of the ongoing work, and ensuring our colleagues in LAs and the wider sector had opportunities to discuss the project with us directly.

We also shared drafts of our core deliverables as we arrived at them, including early publication of the framework for presenting and maintaining the SSD specification. As such, each of our products has more than one "delivery date", and often several; **the below timeline notes the key points at which DfE first received different project outputs**:

May 2023	"Summarising User Research" document and presentation
Jul 2023	First public SSD specification diagrams and production pipeline
Oct 2023	Updated SSD specification diagrams following data mapping
Dec 2024	"SSD New Data Workshop Synthesis" document and presentation
Feb 2024	"Future Support Plan" document and presentation
Mar 2024	Updated SSD specification diagrams following pilot deployments
Mar 2024	First complete data extract production methods
Mar 2024	Final report (this document) and presentation

As our project approached its funded deadline of March 2024, we engaged with DfE about future support plans including producing the above-noted "Future Support Plan" document. At our final steering group meeting of the year, **we agreed to continue meeting as a steering group** into the following financial year, and for D2I to continue working to deploy and develop the product for as long as possible to allow DfE time to consider appropriate arrangements for its continued post-implementation support.

The final month of the project's work largely comprised data validation and testing of our core codebase, making financial arrangements for the end of the funded year, and writing this report.

6. The current product

The current technical product comprises two broad elements:

- The SSD specification (live)
- Generic production methods for generating a local SSD from market-leading CMSs (live)

Using these resources, a competent SQL developer can deploy a functional SSD extract from either of the two market leading CMS products, for use in local reporting, or for potential future participation in standardised data analysis and/or sharing activities.

The wider project output comprises a set of detailed reports arising from the work:

- "Summarising User Research" (June 2023)
- "SSD New Data Workshop Synthesis" (December 2023)
- "Future Support Plan (February 2024)
- "Final Report" (March 2024, this document)

To avoid confusion, it's important to clarify that the project output does not include any new data sharing arrangements. The product is designed to produce a dataset into a local data store, to a national standard. This enables better data collaboration, but does not mandate it.

6.1. The SSD specification

The SSD specification is a dataset specification comprising around 35 separate data tables, their expected contents, and their relationships with each other.

This new dataset specification represents a difference in approach from other existing standard datasets in the sector. Where others are designed to satisfy statutory reporting requirements with the minimum possible number of distinct data items, the SSD is instead designed to satisfy typical local reporting requirements with a suitable balance between completeness, ease of use, and adaptability to different purposes. As such, there are more data items here than in any one of the existing statutory data returns, but for the most part the data transformations required to produce the SSD items from a local CMS are much less complicated.

We don't discuss the specification in detail in this report, because the specification is available as a live public document, including guidance notes and version history. However, it's worth noting a few design principles and decisions informing the shape of the final product:

Flexibility	The dataset should help LAs interact with data and tailor outputs
Completeness	The dataset should include all commonly used items by good LAs
Ease of use	The dataset should make sense to new and learning users
Meeting local needs	The dataset should accommodate local reporting practice
Extensibility	The dataset should be changeable as national data needs change

From these principles we arrive at, for example:

- Normalisation of data into multiple tables – to better match local system shapes, and enable flexible dataset use, as opposed to current statutory reporting which favours fewer tables.
- Extension of data beyond the statutory – to include additional locally-valuable data items where they are "standard enough" across LAs, and broaden the scope for collaboration.
- Full documentation in one place – so that users can easily find and review all available guidance on how to produce the dataset, and what each data item represents.
- Inclusion of multiple years of historical data – because unlike current statutory reporting, local reporting will often consider trends extending beyond the current reporting year.
- Open version control – so that users can benefit from introduction of changes to the specification over time, and collaborate on emerging data requirements.

We see the SSD specification as a living document; like the statutory returns maintained by DfE, and the Annex A specification maintained by Ofsted (but without currently holding the statutory backing which those specifications hold), the SSD will need to adapt over time to remain in step with recording practice in LAs and reporting requirements from government.

Code and guidance for the specification is located at the Data to Insight code repository, hosted by GitHub: GitHub - data-to-insight/ssd-data-model: Standard Safeguarding Dataset (SSD)

6.2. The SSD specification pipeline

In technical terms, the SSD specification is a live, public, open-source document backed by an automated process for generating the various presentation views and associated documentation.

This means that changes to the specification are visible through clear version control and historical documentation, and there is an efficient process for recreating necessary documentation after any change.

The solution is Python based, generating YAML file structures, drawing from a Microsoft Excel working specification document (imported as .csv). This provides flexibility to dataset designers, and reproducibility for documentation managers. It provides an open code base for users to scrutinise as they see fit, paired with straightforward and comprehensive specification documents.

The project has opted for YAML data structures over such as XML for SSD data object definitions. YAML's more human-readable format and provision for annotations directly within the file, is expected to lower the learning curve for LA involvement/further development, enabling both technical and non-technical colleagues to better understand the data object specifications. Concurrently the YAML files have facilitated a more modular approach to configuration, where each data object is defined in its own file, reducing potential overheads for maintaining, extending and scaling the project.

The specification workflow is currently based around the Excel/csv working specification document, however, this workflow stage can be easily removed, and the specification maintained entirely from YAML files within the cloud based environment, should that suit future maintenance arrangements.

The live specification and tools enabling continuous integration/change management is hosted for user navigation on GitHub Pages: data-to-insight.github.io/ssd-data-model/

6.3. The SSD production methods

We now have SSD production methods for the standard LiquidLogic and Mosaic implementations, covering the majority CMS market share in England, and we have tested these across 6 LAs to ensure a base level of compatibility across LA instances. This includes converting the Mosaic code to Oracle-SQL in addition to the original T-SQL approach.

These production methods are, at the time of writing, stored in a private repository within the data-to-insight (github.com) code library.

These methods are currently limited to the two market-leading CMS data sources, and we know that data practice even across LAs using these two systems does vary; LAs will need some support to adapt these methods into their local contexts. There are then also 3 further CMS providers for which we have not yet completed work on SSD production methods, accounting for approximately 20 LAs.

We developed these methods in partnership with pilot LAs, compensating those LAs for the significant investment of time and expertise which they made in this work.

6.4. The original user research summary

We summarised our first round of user research in a report presentation called "Summarising User Research", originally delivered at our open show-and-tell meetings and then provided to DfE as a supporting document.

The report is colour-coded to separate respondent contributions from the project team's headline findings and further commentary, and contains a valuable overview of the research, including a series of charts derived from survey responses, clearly demonstrating both the breadth of practice across different LAs, and the areas of commonality.

Key findings are summarised in section 5.2 above, and **a separate document "Summarising User Research" (June 2023) is available which provides full detail.**

6.5. The "new items" report

Following our second round of user research, we produced a separate document focused on data items or areas which were not sufficiently common across LAs to be considered for inclusion in the SSD. We delivered this in October 2023.

The report documents the specific data areas which participants highlighted for future data development, as well as discussing common issues with data collections and general practice in the sector.

Key findings are summarised in section 5.2 above, and **the full report "SSD New Data Workshop Synthesis" (October 2023) is available which provides full detail.**

6.6. The "future support plan" report

As the project year drew to a close we reviewed our work to date and the new data capabilities we had developed, with a view to their long term support. We identified a way to keep product support in place for the immediate future, and drew up proposals for further support, including completion of some outstanding work items as detailed in section 8.1 of this report.

The full report "Future Support Plan (February 2024) is available in draft form, pending further comment from DfE.

6.7. The final summary report

We wrote our **"DDSF1a final report" (March 2024, this document)**, as a record of the project lifecycle, and a signposting document to the other important project artifacts.

This is an important reference document for colleagues interested in working with the SSD or understanding its design, and we will look to publish this alongside other project outputs when DfE colleagues have considered them.

7. Connections to other projects

7.1. The DDSF 1b projects (i), (ii), (iii), and (iv)

The DfE's original vision for the Data and Digital Solutions Fund (DDSF) was that four data development projects – the "1b" projects – would produce outputs which could be assimilated into the SSD on completion. These were envisaged as projects to develop new standard datasets for key areas of interest – **kinship care, pre-proceedings, Section 251 (financial information), and the voice of the child**.

One of these projects – pre-proceedings – produced an indicative dataset and completed a test collection from a large number of LAs, before handing over project outputs to DfE. The new dataset has not yet been adopted for sector-wide use. In the SSD, we have built the likely data format into our specification, but have marked this as currently inactive, as LAs are not currently commonly gathering this information in their core CMS products.

Of the remaining projects, the S251 and child voice projects are likely to produce similar recommendations to DfE, either for new data items in a proposed structure, or for reorganisation of existing data into new structures. At the time of writing this work is incomplete.

The kinship care project became an investigation into the scale of kinship care nationally, rather than a data development project.

In each case, our mitigation against later developments in these projects is in the extensible, flexible nature of the specification we have produced, and the automated pipeline for producing specification documents in light of structural changes. Our approach to shared data extract code will also minimise the work of extracting new data items from local CMS products – but will not eliminate it. In order to align these projects, **new data requirements of LAs (or of CMS products) should consider the SSD during their development**, and future support for the SSD should include provision for adapting to new data needs.

7.2. The DDSF 2a projects

The "2a" projects, led respectively by a partnership of Social Finance and Leeds City Council, and by a partnership of Essex, Camden, Croydon, and Sutton Councils, performed detailed **research into social workers' experiences of using CMS products**. The projects identified the value of different data items recorded by social workers, and the reasons for which different data were collected.

Each project arrived at full reports to DfE including a set of recommendations for further work in this space. **Should DfE look to revise statutory collections in light of these reports, the SSD should adapt in light of these changes**, and the SSD should also be considered as a possible venue for data sharing between LAs and DfE.

7.3. The RIIA Quarterly Dataset

LAs collaborate on a quarterly aggregate data collection called the RIIA Quarterly Dataset, to provide **in-year benchmarking and early trend analysis** across a suite of high-level performance measures.

The SSD should, as our work to support it develops, be capable of producing the bulk of the RIIA dataset for LAs without bespoke local data work.

The SSD may, if adopted across the sector, increase the potential scope of the RIIA collection; the exercise is limited in order to keep its outputs timely for leaders, and low impact for data officers who must produce the dataset alongside other duties. Should the SSD succeed in facilitating broader standard data output from LAs, it could be used to produce a broader dataset on behalf of LAs to support their in-year analysis.

7.4. The DfE CIN Census and SSDA903

Should DfE look to revise statutory collections in light of the above-discussed data development projects, the SSD should adapt in light of these changes, and the SSD should also be considered as a possible venue for data sharing between LAs and DfE.

The CIN Census and SSDA903 collection have been DfE's primary mechanism for understanding children's safeguarding data for around two decades; they are long-established processes well-integrated in local business processes, CMS products, and national policy work. DfE's new data strategy for children's social care may result in significant change to these collections, and **the SSD and its project partner organisations are well-placed to support any such change**.

7.5. The Data to Insight sector-led service

Data to Insight (D2I) is the sector-led service for children's safeguarding data professionals. Maintaining sector-critical data tools, data collections, and learning venues, D2I provided the technical team for the SSD project, as well as convening the user research phases of the work, and supporting the steering group. D2I has been a crucial support venue for data-focused collaboration in the sector, and was recognised by name in the recent MacAlister review of children's social care, with the review recommending that DfE "should work closely with Data to Insight".

The SSD will rely on D2I for its short-term future support in 2024-25, while DfE considers the project outputs and the recommendations in our separate "Future Support Plan" report. SSD and related work should also consider D2I a key partner for helping make best use of standard datasets, as well as for supporting LA analysts and the wider sector data community in working together on shared approaches to common challenges.

D2I is a small, low-cost team, hosted by one LA, and led by a board of volunteer directors representing other LAs, regional networks, and other sector leaders. To ensure full accessibility of its offer to every LA, it currently draws most of its funding from competitive central government opportunities.

As part of furthering the good work done on recent data innovation projects, **DfE should seek a sustainable funding approach for national sector-led data collaboration, including D2I**, which can effectively support the existing value in the sector.

7.6. Other local government use cases

This is a non-comprehensive, indicative list of **example use cases for the SSD** once implemented, to help stakeholders understand the value of the approach and its potential to support cross-sector goals. Though we've split this between LA-centric and DfE-centric use cases, in most cases, anything improving data work in the sector will ultimately benefit both DfE and LAs.

- Extended ChAT – potential to make shared tools like the ChAT more useful
- Patch – potential to extend the D2I "patch" tool with wide-ranging online analysis tools
- FOI response work – potential to share scripts for answering FOIs based on standard dataset
- Demand modelling – potential to make shared demand tools more insightful with wider standard dataset available for modelling
- Responding to DfE data requests – potential to share scripts for answering DfE requests
- Quarterly benchmarking – potential to dramatically reduce admin burden of quarterly returns
- Quarterly benchmarking – potentially to dramatically expand reach of quarterly benchmarking without increasing admin burden for individual LAs
- Further collaboration – with our away from D2I, standard data formats enable collaborative projects between LAs nationally to improve processes, better understand markets, etc.

7.7. Other central government use cases

This is a non-comprehensive, indicative list of **example use cases for the SSD** once implemented, to help stakeholders understand the value of the approach and its potential to support cross-sector goals. Though we've split this between LA-centric and DfE-centric use cases, in most cases, anything improving data work in the sector will ultimately benefit both DfE and LAs.

- Aligning with uses of DfE's proposed Common Child Identifier (CCI) and other systems-connecting initiatives - we have future-proofed the SSD specification by including fields for this in the draft specification, and for other relevant identifiers
- Ad hoc data requests to LAs – potential to provide SSD extract script alongside request to expedite LA responses
- Improvement of statutory returns – potential to use SSD as a combined stat return with higher information density, or to select SSD items for inclusion in statutory returns
- Data flows – potential to use SSD as the LA-DfE interface for any higher frequency data sharing between LAs and DfE
- Quarterly benchmarking – potentially to dramatically expand reach of quarterly benchmarking without increasing admin burden for individual LAs
- Responding to DfE data requests – potential to share scripts for answering DfE requests
- Improving CMS – potential to use SSD as standard output layer for a new modular CMS specification
- PLO/pre-proceedings – potential to incorporate to SSD for shared national analysis approaches
- Other data development – potential to reduce implementation overheads of new data requirements for LAs (notwithstanding CMS supplier costs would remain)

8. Benefits realisation / measuring impact / measures of success

We endorse the DfE vision of a standard safeguarding dataset in use by every LA to improve data collaboration within the sector and between LAs and DfE. As such, the most obvious questions to ask of a project like this are "how many LAs are using it?" and "what good is it doing?"

These measures are outside the scope of the current project, which concludes with the delivery of the specification and the methods for producing the dataset. We look forward to working with DfE and colleagues in LAs to design the next stage of work which should answer these questions.

In the meantime, we'll measure success of the current project in terms of the original project brief.

8.1. Satisfaction of the project brief

The brief called for two key deliverables, preceded by a first phase of research, as detailed in section 4 above.

Stage 1:		
Data mapping	✓	Data mapping completed including consultation
User research	✓	Two phases of research completed successfully
Research report	✓	Two reports provided to DfE
Dataset specification	✓	Live online specification available to all stakeholders
Comparison to existing returns	✓	Detailed in live online specification
Stage 2:		
Production methods	📈	Production methods for 2 market-leading CMSs
Maintenance framework	✓	Framework live and documented via GitHub

We've detailed our approaches to each of these deliverables in earlier report sections. The final two are those requiring immediate further work: most pressingly, there are three CMS products in use across around 20 LAs for which we do not yet have production methods for the SSD.

This represents less than 15% of LAs, but **any standard is only as strong as its market share**, and we believe that it's crucial to produce shareable code for these remaining LAs even though the per-LA cost is necessarily higher. Our project was unable to complete this work according to the original schedule, in large part because, with a smaller user base, we were unable to find partner LAs for these systems with suitable technical resource available at short notice.

In each case we have identified a partner LA and begun working with them on a delivery plan, but these plans will extend beyond the current financial year and are therefore currently unfunded. We are working with DfE to ensure this work is completed in the next financial year.

The final deliverable – the maintenance framework – is complete as a technical product, but incomplete with regard to future support and governance arrangements. **We have agreed to provide basic support through D2I for the next financial year**, and agreed to continue meeting as a steering group for the same period, to facilitate further work on these challenges. More information on this work can be found in the separate "Future Support Plan" document.

All outputs were delivered by March 2024. With the above caveats, we consider this a successful delivery on the original project brief.

8.2. Cost

We spent approximately £400,000 less than initially anticipated in delivering this project. This was achieved primarily through identifying more effective delivery routes than initially anticipated.

While we recognise above that approximately £100,000 of work remains unfinished from the original brief, this represents a major achievement on the part of the partnership to adapt appropriately during delivery, and to identify better ways of working than we originally envisaged.

Including the **commitment from D2I and the steering group to continue supporting the project into the next year**, so as to ensure time for DfE to consider the outputs and their future value to the sector, we believe this project represents good value for money.

8.3. Feedback

We also received feedback from project colleagues, research participants, open meeting attendees and DfE colleagues which indicated broad approval of our work, both in how we ran the project and in our project outputs. We note below a selection of comments from our project's final open "show-and-tell" meeting, attended by steering group members, project team officers, DfE stakeholders and colleagues from other LAs:

> "This is **really valuable work**, and it has been really useful to have worked in such an open and collaborative way."

> "It feels like a great cross-LA project to benefit LAs."

> "**Really well delivered**."

> "Thank you to the team at D2I for making this all happen – and to our LA partners, without whom it wouldn't have been possible."

This feedback, and other instances much like it through the longer history of the project, aligns with our own experiences. We have seen colleagues in LAs, in the core project team, and in the supporting DfE teams, working well together, and effectively contributing their effort and expertise.

8.4. Future work and associated potential impact measures

As noted above, **the most obvious performance questions for the SSD are "how many LAs are using it?" and "what good is it doing?"** Should a future project phase emerge to support these products, we would add to these:

- Partnership engagement measures, to ensure LAs are represented in future work
- Local data burdens measures, to understand where we are enabling better practice
- Local data impact measures, to understand which data practices best support safeguarding

However, we know that **the throughline from data practice to service improvement is difficult to measure quantitively**. Too many confounding factors make it difficult to quantify the impact which improved data practice might have had on any individual LA's safeguarding performance. As such, we would recommend taking the above as indications of what we would like to understand, rather than a framework for devising specific performance metrics.

We believe there is much future work to do in support of this project's outputs, to ensure that our work can best benefit the sector. We detail this in the separate "Future Support Plan" report.

9. Learning from the project

Through our work we gained valuable insights into data practice across the sector, which warrant the separate detailed research reports we made and discussed in sections 5, 6.4, and 6.5 above.

Here, we offer our reflections on the project as it ran, and the learning we would want to capture and share for the benefit of other similar projects and initiatives.

9.1. Stakeholder engagement

We believe in partnership working, and in open working, as effective tools for delivery in this sector. We actively sought input from all potential stakeholders, and directly engaged with a broad spectrum of roles.

The willingness of those within the sector to engage with, and contribute experience to, the project has been critical to refining the project's core goals and validating our approach. The cross-sector vantage point of D2I particularly helped us to gather and understand high quality sector knowledge and experience contributions from a broad cross-sector community.

9.2. Shared and remote testing

As the project moved through development, testing and pilot deployment, we encountered familiar bottlenecks resulting from justifiable restrictions in LAs' data infrastructure and systems access practices. These required us to develop in distinct workstreams and to create products which operated solely within individual LAs – though the deployment code can of course be shared and reproduced.

Our response to these challenges was to recruit partner LAs to work with our project team, deploying our code and reporting testing results, or developing their own code and then sharing it with us, without our directly accessing their systems. This approach worked, but was made significantly easier by partnering with D2I's host LA for the first pilot development, where our project team *could* have direct access to code and systems, and thereby learn before working with external partners.

Indeed, concern remains in some LAs regarding the over-sharing of information about their CMS or data. Less data mature LAs will typically look exclusively to their CMS supplier for guidance on data and systems work, which may further entrench a siloed culture and hinder compatibility and interoperability work – although our wider research suggested that these were key aspirations for many LAs.

9.3. Feedback driven iteration

Working in the open reassured stakeholders as to project progress and direction; partners and wider community members alike could review our progress as it happened and provide iterative feedback – upon which we acted.

This will continue into the next development phase, and feed directly into the project's development backlog/bug ticket system. This process has worked well so far, and an increasing level of direct feedback is anticipated as more LAs have hands-on experience with the SSD structure and tools, and can better engage with beta data testing. Feedback to date has been insightful, and directly correlated to periods directly following project presentations.

Given the known time constraints and varied skills-confidence within the sector, it was effective to ensure a persistent and minimalistic channel for wider feedback.

9.4. Direct stakeholder driven iteration

Gaining support from a project 'champion' within a given LA improved the development progress; the project benefited directly from having senior colleagues with direct access to LAs' data/infrastructure teams, in order that more simplistic database system tests and sample runs could be achieved with very low time-overheads. This enabled an accelerated development cycle with specific pilot partners.

Our iterative approach also allowed us to overcome unexpected hurdles – for example, learning mid-project that some CMS platforms can operate on more than one database technology, dependent on LA preference. This naturally complicated development work, but the stakeholder-driven iterative approach helped us find issues early and adapt our work to address them.

9.5. Trust dependent rapid development

We were able to improve development pace where our tests could be run by colleagues without elevated privileges on the required systems. Direct feedback of this sort enabled a rapid iterative development cycle, and this process has been heavily dependent on the trust from LAs towards D2I, and a strong desire on the part of LAs to support positive change in the sector.

We also saw evidence that increased development and testing pace was directly tied to levels of consistency across deployed instances of each CMS; CMS products allow for different kinds of customisation, with some products enabling higher levels of database structure flexibility to customers. This alters the pathway for managing standard data outputs.

9.6. Proven basis for collaborative tool development

The project was able to prove the D2I-championed concept of "Build once, use 150 times" for collaborative data work. Tools and scripts written by one LA can be deployed in other LAs, and a small central resource can facilitate valuable sector-led improvement.

Scripts for some lists were developed on one system(T-SQL+LiquidLogic) and others concurrently developed on another (T-SQL+Mosaic). The combination of this effort, and any further developments can now be utilised by any LA.

9.7. Scale as a barrier

Our project set out to produce data extract code for five CMS products, and did so for two. A standard is only as valuable as its acceptance rate, and so we hope to continue this work for the remaining three products into the coming year.

With smaller market share, these three products offer us fewer potential partners for pilot work, requiring us to work at our partners' pace, not our own, if we wanted to deliver. We found similar learning in our user research: LAs are all different, and different kinds of LA will have different needs, aspirations, and issues with any new work. Some will struggle to contribute time to innovation projects, while others are well equipped to lead the work; but both must be well served by the output if it is to generate real value. A key learning point is that we must consider the full breadth of LA contexts in data innovation work.

9.8. DfE's role in sector data development

Our work discovered and demonstrated successful instances of sector-led data development, and we engaged with many LAs working independently on their own data development agenda. We also recognised widespread reticence to adopt data practice or production burdens where DfE did not require it. Some LAs cited limited resources, or increasing data burdens, or CMS costs and limitations, as reasons for deferring on voluntary or sector-led data opportunities. We also heard from some LAs who were unwilling to engage in new data activity unless funding provision was made for that new activity.

From this we conclude that the SSD's success as a national standard may be contingent on DfE's adoption of the standard as a national requirement. While initial interest from LAs in helping us test the SSD output code is very encouraging, we would not expect 100% of LAs to volunteer to adopt the SSD production methods unless, as with Ofsted's "Annex A" or DfE's annual data collections, some kind of national framework exists to make this a necessary part of LAs' data work.

10. Conclusion

The project has now delivered on most of its original scope at significantly lower cost than originally expected. However, DfE funding restrictions mean we are unable to "carry" funding for the remaining project strands into the next financial year. As such, **our primary consideration at the time of writing is to keep the project's products in active support** while DfE considers their longer-term future.

We also seek to maximise uptake of the SSD in LAs, because we know that the usefulness of any standard is directly proportional to its universality; if not all LAs can produce data to the standard, this severely limits its usefulness to adjacent workstreams and future initiatives.

At the close of the initial, we recommended a twofold approach to meeting these requirements in the short term, both of which we have now agreed with DfE.

First, we recommended using D2I to continue to support roll-out and maintenance of the SSD specification and production methods through the next financial year.

Second, we recommended a new funding agreement with DfE, to be paid in the next financial year, to complete development of extract methods for the three small-market-share CMS products.

Further to these initial recommendations, we also recommend that DfE incorporate in its CSC data strategy a provision for long term sector support of initiatives like this, including the SSD and other sector-led data work. This support should be adaptable to the sector data landscape as it changes, ensuring the SSD a place in the sector's digital and data landscape beyond the current support roadmap, which ends in March 2025.

As detailed above, we are delighted to have been part of bringing this project to fruition, and **we believe the project has delivered a product far surpassing initial expectations, and at a lower cost than originally anticipated**. We look forward to seeing further work in this area from colleagues across the sector to make best use of the new capabilities of the SSD.

11. Acknowledgements

This project would not have been possible without the generous and insightful contributions of a great number of colleagues across the sector, or without the dedicated support of our steering group members and partners.

We are particularly grateful to our **steering group**, made up of colleagues from:

> Data to Insight, hosted by East Sussex County Council
> Essex County Council
> Hertfordshire County Council
> Knowsley Council
> North-West Association of Directors of Children's Services

We also extend thanks to our **wider circle of contributing LAs**, especially those which endorsed our project bid, or later assisted with developing or testing our research and technical products:

> Blackpool Council
> City of Wolverhampton Council
> Dorset Council
> East Riding of Yorkshire Council
> East Sussex County Council
> London Innovation and Improvement Alliance
> Manchester City Council
> Newham Council
> Staffordshire County Council
> Sutton Council
> Trafford Council
> Waltham Forest Borough Council

Further, we are grateful to the following **partner organisations** for key contributions to the project:

> Association of Directors of Children's Services
> Coram Voice
> Department for Education

LiquidLogic/System C
Ofsted
University of Sussex

Finally, **we are indebted to around 100 other colleagues from LAs and other organisations** who contributed to our anonymous research exercises. Without this work, the project would have been immeasurably poorer.

12.   Appendix 1: User Research Synthesis

A separate report detailing the technical workshops we held during the project is available from the website www.datatoinsight.org/ddsf

13.   Appendix 2: Technical Workshops Synthesis

A separate report detailing the technical workshops we held during the project is available from the website www.datatoinsight.org/ddsf

To further discuss the content of this report, please contact datatoinsight.enquiries@gmail.com