

Article

How Can a Clinical Data Modelling Tool Be Used to Represent Data Items of Relevance to Paediatric Clinical Trials? Learning from the Conect4children (c4c) Consortium

Chima Amadi ¹, Rebecca Leary ¹ , Avril Palmeri ¹ , Victoria Hedley ¹, Anando Sen ¹ , Rahil Qamar Siddiqui ², Dipak Kalra ³ and Volker Straub ^{1,*}

¹ John Walton Muscular Dystrophy Research Centre, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE1 3BZ, UK; c.m.amadi2@newcastle.ac.uk (C.A.); becca.leary@newcastle.ac.uk (R.L.); avril.palmeri@newcastle.ac.uk (A.P.); victoria.hedley@newcastle.ac.uk (V.H.); anando.sen@newcastle.ac.uk (A.S.)

² Sidqam Ltd., Altrincham WA14 3NB, UK; rahil@sidqam.co.uk

³ Department of Medical Informatics & Statistics, The European Institute for Innovation through Health Data, Ghent University Hospital, 9000 Gent, Belgium; dipak.kalra@i-hd.eu

* Correspondence: volker.straub@newcastle.ac.uk

Featured Application: To demonstrate the advantages of using a modelling tool to define and structure the clinical data items represented in the conect4children (c4c) Cross Cutting Paediatric Data Dictionary (CCPDD). By facilitating the unambiguous definition of data items, we show how a modelling tool can improve the interoperability of data collected in paediatric clinical trials. We demonstrate how the clinical application of a modelling approach to semantic interoperability can be used to represent the c4c CCPDD. We highlight how modelling via tooling of terminology is a better means of data dictionary representation compared to an implicit model captured in Microsoft Excel. We report on how the Direcht tool facilitates the export of data models into electronic capture (EDC) systems, such as REDCap and Castor. Finally, we illustrate how the Direcht tool formalises the modelling of data items to improve the content of the CCPDD and makes it semantically and contextually complete.

Abstract: Data dictionaries for clinical trials are often created manually, with data structures and controlled vocabularies specific for a trial or family of trials within a sponsor's portfolio. Microsoft Excel is commonly used to capture the representation of data dictionary items but has limited functionality for this purpose. The conect4children (c4c) network is piloting the Direcht clinical data modelling tool to model their Cross Cutting Paediatric Data Dictionary (CCPDD) in a more formalised way. The first pilot had the key objective of testing whether a clinical data modelling tool could be used to represent data items from the CCPDD. The key objective of the second pilot is to establish whether a small team with little or no experience of clinical data modelling can use Direcht to expand the CCPDD. Clinical modelling is the process of structuring clinical data so it can be understood by computer systems and humans. The model contains all of the elements that are needed to define the data item. Results from the pilots show that Direcht creates a structured environment to build data items into models that fit into the larger CCPDD. Models can be represented as an HTML document, mind map, or exported in various formats for import into a computer system. Challenges identified over the course of both pilots are being addressed with c4c partners and external stakeholders.

Keywords: conect4children; clinical modelling; data dictionary; paediatric clinical trials; electronic data capture; Direcht; Castor; REDCap; CDISC



Citation: Amadi, C.; Leary, R.; Palmeri, A.; Hedley, V.; Sen, A.; Siddiqui, R.Q.; Kalra, D.; Straub, V. How Can a Clinical Data Modelling Tool Be Used to Represent Data Items of Relevance to Paediatric Clinical Trials? Learning from the Conect4children (c4c) Consortium. *Appl. Sci.* **2022**, *12*, 1604. <https://doi.org/10.3390/app12031604>

Academic Editor: Andrea Prati

Received: 13 December 2021

Accepted: 28 January 2022

Published: 2 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Data obtained from vulnerable populations, such as babies, children, and young people, is precious. When seeking to optimise the design and delivery of paediatric

trials, data considerations are paramount. At present, the lack of harmonisation and standardisation concerning the collection and representation of data elements in paediatric clinical trials is a significant challenge. *conect4children* (c4c) is a six-year project established in 2018 and funded by the Innovative Medicines Initiative (IMI2) to address the multitude of barriers facing the development and delivery of effective paediatric clinical trials. This will include a trial feasibility, conducted through a single point of contact, expert advice, an educational academy for clinical trial personnel and data harmonisation services [1–3].

Clinical trials that contribute to drug registration are tightly regulated. Regulation includes the use of explicit data structures and controlled vocabularies. At present, many electronic case report form (eCRF) specifications are created manually with data structures and controlled vocabularies that are unique to each clinical trial, including those in paediatrics. That is, each trial (or family of trials within a sponsor's portfolio) has an implicit model, is designed for an ad hoc purpose, and is limited by how the model is captured (usually in a flat file or spreadsheet). There is presently a limited culture and practice of standardising and reusing clinical data specifications between clinical trials, and no vendor neutral specification for automated eCRF generation. The Clinical Data Interchange Standards Consortium (CDISC) standards—in particular, its Therapeutic Area User Guides (TAUG)—support access to data by specifying how data standards are expressed. The CDISC also encourages convergence in data standards across pharma clinical trials [4]. There is presently limited connection between the CDISC standards and those used for routinely collected (real-world) electronic health records, although the HL7 VULCAN project is starting to bridge between its Fast Healthcare Interoperability Resources (FHIR) standard suite and clinical research [5]. However, there is a growing recognition that clinical trial data should be reused for subsequent research [6], as exemplified by the Yale Open Data (YODA) initiative supported at Yale University by Janssen Pharmaceuticals [7]. Eligibility criteria for clinical trials are already being mapped into a computable form for execution as queries on hospital electronic health records (EHR) systems that requires reformulating those criteria as data items and structures that are realistic to find in EHRs [8,9]. More recently, vendor neutral specifications for transferring hospital EHR data to electronic data capture systems (EDC) for pharma-sponsored trials has been demonstrated at multiple European sites [10]. It is therefore important to find ways to maximise the consistency and reusability of clinical trial data.

The c4c project is addressing this issue by developing and supporting a standardised paediatric data dictionary, the Cross Cutting Paediatric Data Dictionary (CCPDD). This provides guidance on how to represent data items that are disease agnostic (cross cutting) and commonly collected during paediatric studies. Use of the CCPDD is expected to result in more harmonised, interoperable data that has a greater potential to be shared and re-used. The CCPDD has an explicit model and is designed to be used across multiple trials and sponsors.

The c4c CCPDD was initially developed in Microsoft Excel and was successfully used to manually create case report forms (CRFs) for the non-industry c4c Proof of Viability (PoV) studies. However, the Excel format of the data dictionary resulted in several challenges due to the 'flat' structure of the worksheet and its limited ability to quality assure the content entered.

Due to all the issues listed in Table 1, the c4c Paediatric Data Interoperability Working Group (PDIWG) strongly recommended future versions of the CCPDD use a more formalised modelling approach to defining the data structure.

Table 1. Challenges of using a Microsoft Excel spreadsheet to represent the CCPDD.

Issue	Explanation
Version control	As the Cross Cutting Paediatric Data Dictionary (CCPDD) grew and new versions were developed, there was a risk that obsolete versions of it were used. The size of the connect4children (c4c) consortium made this more likely.
Complexity	As the CCPDD grew, it became more cumbersome to manage. Adding new worksheets to the spreadsheet made the document less user-friendly.
Modelling errors	Risk of data dictionary concepts being modelled differently in different domains by different teams.
Linkage	Due to the flat structure of Microsoft Excel, it was challenging to show nested structures and interlinking items.
Semantic detail	Not all structure and semantic specifications can be captured, and there was no way of prompting authors about missing data properties.
Usability	As the CCPDD grew, visualisation became increasingly important to help users to understand what the dictionary contains, and what they were now in the process of authoring. This functionality was not possible in Microsoft Excel.
Only one standard used (CDISC)	Predominantly focussed on one standard (Clinical Data Interchange Standards Consortium (CDISC)): this reduced opportunities to link items to data from electronic health records.
Integration to other IT tools	The Microsoft Excel version required transcription into electronic case report form (eCRF)-generating tools.

In contrast to clinical research, methods and tools to formalise clinical models (data structures and semantics) that define documentation patterns for the consistent population and interoperable communication of EHRs are well established [11].

The openEHR archetype concept is the oldest and most mature formalism for this [12,13]. In more recent years, other initiatives have adopted similar approaches [14], and quality processes for developing clinical models have also been published [15,16]. This clinical modelling paradigm has been an ISO standard since 2008—most recently published as ISO 13606 Part 2 [17]—which enables clinical models (archetypes) to be defined as use patterns of an underlying semantics-free reference model defined in ISO 13606-1 [18]. Clinical modelling tools have been used extensively by EHR interoperability communities globally in order to help them to define models that are technically and semantically aligned with an underlying reference model and offer a number of user features to facilitate good quality model development [19]. Mapping transformations have been defined between archetypes based on ISO 13606 and the CDISC Operational Data Model (ODM) standard most often used for clinical trials data [20]. There remains, however, an evidence gap in terms of the incorporation of the archetype approach within regulated clinical trials by leveraging this computable knowledge representation as a source for eCRF generation and CDISC ODM mapping. c4c therefore elected to pilot the use of the archetype approach—based on the

ISO 13606 standard—to reformulate and to better formalise its initial CCPDD, including the export of these paediatric research archetypes into a CDISC ODM format and the automated configuration of eCRF templates.

The scope of the problem under investigation was whether it was possible to robustly and methodically structure the content of the CCPDD into an environment that was not compliant with clinical trial standards alone. The reasons for exploring this were twofold:

- (1) to enable data from the clinical trial environment to flow seamlessly into point of care health environments;
- (2) to validate and improve the content within the data dictionary, to make it semantically and contextually complete.

A data dictionary modelling pilot project was established to reproduce selected parts of the c4c CCPDD using the Direcht modelling tool (<https://direcht.com/> (accessed on 20 January 2022)). The Direcht tool has been developed by a team with a long history in the clinical modelling field within the health domain, including serving as ISO experts for the development of the 13606 standard. Direcht conforms to this ISO 13606–2 clinical modelling formalism, as well as the 13606–1 underlying reference model. Because these two-part standards themselves do not incorporate clinical model content and can be transformed into other clinical and research representations, the c4c team regarded this approach as standards and semantics neutral with respect to the more specific implementation standards that are used in healthcare and in clinical research.

The data modelling tool, as it stands, thus provides a standards-agnostic—but nonetheless standardised—data dictionary representation, which supports alternative export formats, such as those used in the wider healthcare domain.

The ability to represent the dictionary in a three-dimensional tool using the CDISC ODM was important as this would allow the content of the dictionary to be imported into one or more EDC systems.

Recent research conducted with the c4c industry partners showed that there was interest within industry for c4c developing a resource where data are modelled and standardised to avoid duplication of effort and allow for greater harmonisation.

2. Materials and Methods

The Direcht modelling tool [21] utilises various modelling paradigms from the ISO 13606–1 Reference Model [22,23], such as ‘Element’, ‘Cluster’, ‘Entry’, and ‘Sections’. For instance, ‘BodyMeasurement’ is a ‘Cluster’ which encompasses different data dictionary items such as ‘BMI’ and ‘height’ (represented as ‘Elements’). These ‘Clusters’ and ‘Elements’ can be brought together under an ‘Entry’ that represents a well-defined, communicable piece of information. One or more of such ‘Entries’ can be grouped together under headings, represented as a ‘Section’ within the tool.

Figure 1 shows the procedure used for the modelling of data dictionary items in Excel, a snapshot of which is shown in Figure 2. BodyMeasurement (Vital Signs) was used for the first pilot study using Direcht and reproduced at the start of the second pilot. The first step was the creation of an account in Direcht. Modelling of the data dictionary items was started under the project CCPDD–pilot 2 within the ‘Create Clinical Model’ section as shown in Figure 3 below. In the ‘Entry’ section, the page was completed, which included ‘Entry Name’ (‘BodyMeasurement’), ‘Specialization’, ‘Version Status’, ‘Conformance’, ‘Cardinality’, ‘Version Number’, ‘Entry Description’, ‘Mapping Type’, ‘Mapping URL’, ‘Mapping Description’, and ‘Elements’. The ‘Conformance’ and ‘Cardinality’ values were selected on the basis of the importance of the data dictionary item in the specific aspect of a clinical trial. Moreover, it is important to recognise that not all of the sections were compulsory to complete, and as such, compulsory fields were marked with a red asterisk. In the mapping section, the CDISC link to the data dictionary item being modelled was entered in the ‘Mapping URL’ section, the description of the entry was stated in the ‘Mapping Description’ section, and the ‘Mapping Type’ was labelled as ‘CDISC’.

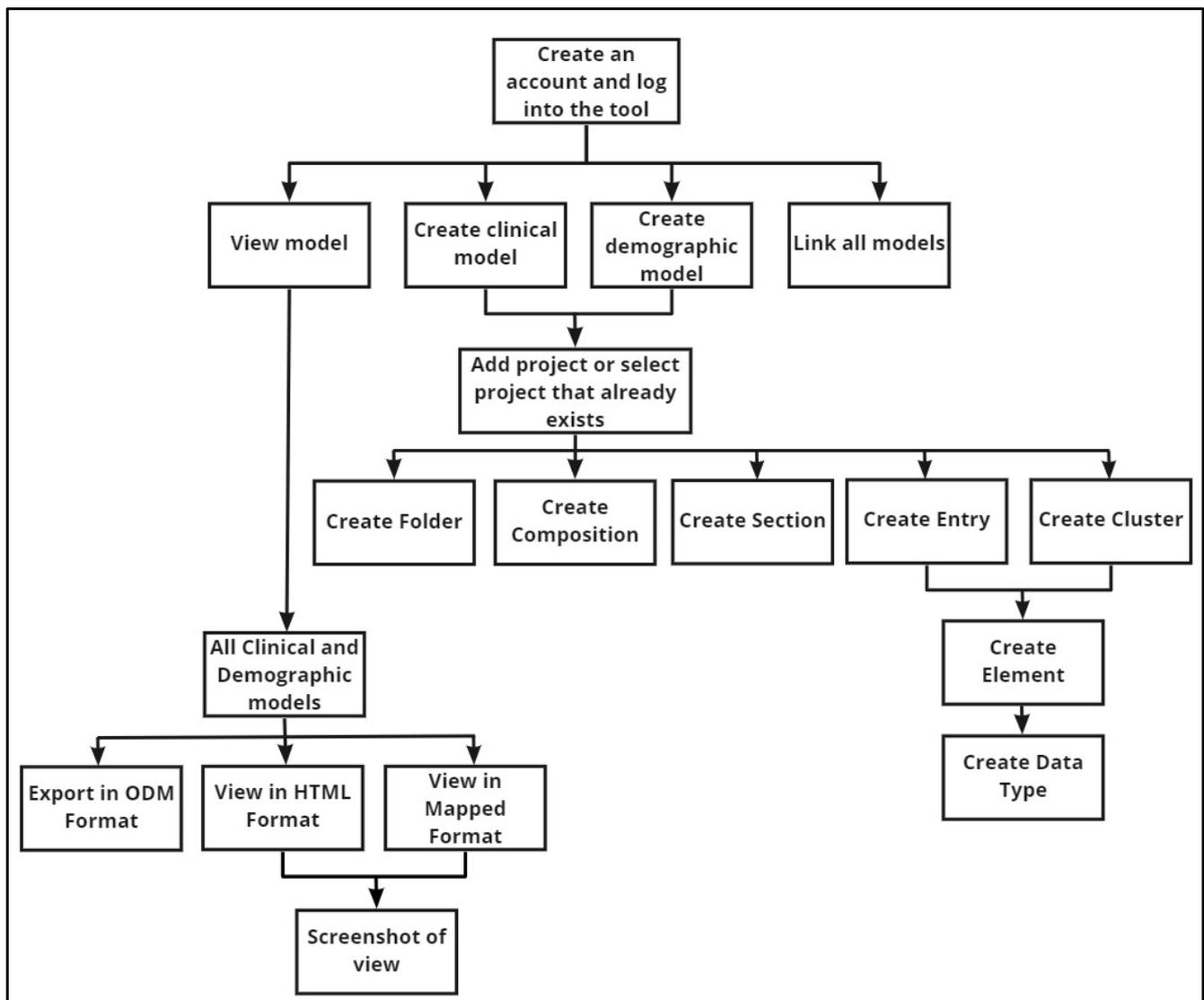


Figure 1. Schematic diagram showing the flow chart of the use of Direct to model data dictionary items. HTML—hypertext markup language; ODM—operational data model.

Through the data modelling tool, one can model data dictionary items in a non-CDISC standard environment. Moreover, data content can be exported to the CDISC ODM format and imported to one or more clinical EDC tools. These features of the tool provide an avenue for data from the clinical trial environment to flow effectively to the point of healthcare where they are needed. It also provides a better structural and visualisation of data for use by those at the lower levels of the clinical trial ladder across EDCs. Furthermore, it helps to make the data dictionary contextually complete.

In the ‘Element’ fields, as shown below in Figure 4, four ‘Elements’ were completed, namely, ‘BMI’, ‘Height’, ‘TotalBodyLength’, and ‘Weight’. The ‘Cardinality’ and ‘Conformance’ of these ‘Elements’ were completed according to the importance of the measurements in specific clinical trials. Following this, the ‘Entry’ was saved.

CDISC mapped Item	CDISC Qualifier Variables	Guidance	Source of information CDASH = Data Collection SDTM = Data Tabulation
Total body Length	Vital Signs Position of Subject (VSPOS)? - not needed if new definition accepted as it states "recumbent"	Specify standardised unit	CDASHIG v2.1 - https://www.cdisc.org/system/files/members/standard/foundational/CDASHIG%20v2.1-Final_Rev.pdf Section 8.3.18 SDTMIG v3.3 - https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-3/html#Vital+Signs
Weight		Specify standardised unit	CDASHIG v2.1 - https://www.cdisc.org/system/files/members/standard/foundational/CDASHIG%20v2.1-Final_Rev.pdf Section 8.3.18 SDTMIG v3.3 - https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-3/html#Vital+Signs

Figure 2. Microsoft Excel worksheet showing the c4c data dictionary item for BodyMeasurement. CDASH—Clinical Data Acquisition Standards Harmonisation; SDTM—Study Data Tabulation Model; CDISC—Clinical Data Interchange Standards Consortium; CDASHIG—Clinical Data Acquisition Standards Harmonisation Implementation Guide; SDTMIG—Study Data Tabulation Model Implementation Guide.

Figure 3. Image showing the ‘Create Entry’ screen in Direcht.

ATTESTATION INFORMATION		AUDIT INFORMATION		OTHER ENTRY ATTRIBUTES	
Domain					
Select Domain					
Relationship	Element Meaning	Element Name *	Cardinality	Conformance	Value Type
Select Relationship			Select Cardin	Select Conforma	Select Value Type
Add Element					
SNo	Select Cluster				
	Select Cluster	Add Cluster			
Author Name		Authoring Date			
Anando Sen		20/01/2022			
Save Entry		Cancel Entry			
Copyright © 2022 Sidqam Limited (Direcht). All rights reserved.			Support: direcht@sidqam.co.uk t:+44 (0) 7448939001		

Figure 4. Image showing the ‘Element’ section in an ‘Entry’ page in Direct.

Afterwards, the ‘Create Cluster’ page was opened and, like the ‘Entry’ page, completed. The ‘Clusters’ were named ‘BloodPressureMeasurement’ and ‘BodySurfaceAreaDetails’. ‘Elements’ completed included ‘SystolicBloodPressure’, ‘DiastolicBloodPressure’ and ‘BodySurfaceArea’. Finally, the data dictionary items modelled in the ‘Clusters’ were then attached to the already modelled ‘Entry’ and saved.

3. Results

Eighteen data concepts were included in the first pilot study—10 clinical data concepts and 8 demographic data concepts. The 10 clinical data concepts were as follows: ‘OxygenSaturation’, ‘HeartRate’, ‘PulseRate’, ‘RespiratoryRate’, ‘Temperature’, ‘MeasurementDetails’, ‘BloodPressureMeasurement’, ‘BloodPressure’, ‘BodyMeasurement’, and ‘VitalSigns’. The eight demographic data concepts were: ‘PatientDetails’, ‘PatientInformation’, ‘GgestationalAgeDetails’, ‘PatientDetails’, ‘DateOfBirth’, ‘DateOfDeath’, ‘GestationalDetails’, and ‘PhenotypicSex’.

3.1. Facilitation of Good Quality Model Design

For the second pilot, a small team with no previous experience using a modelling tool was tasked with modelling additional concepts associated with the CCPDD. The more sophisticated design of the modelling tool compared to the Microsoft Excel spreadsheet prompted discussions about the structure of the models that had previously been defined in Excel and how they would fit as the data dictionary increased in size and complexity. This included discussions about reusability of data items and how to structure models so that they could potentially have a more general application. The team agreed that the modelling tool directly facilitated these conversations about modelling aspects that had been overlooked when using Excel and, to an extent, provided a learning environment for colleagues who had no prior experience with data modelling.

As well as helping to define the structure of the data items, the modelling tool also prompted discussions about the cardinality and conformance of data items which was not addressed in the Excel version of the CCPDD.

As described above, the modelling tool provided a platform for discussion and for reaching consensus around building more complex and reusable data models, but it was recognised that the models also needed to be exportable to EDC systems to work in a

real-world setting. The more structured approach of the modelling tool also highlighted gaps in knowledge which resulted in the pilot team reaching out to colleagues within the c4c consortium for guidance on how to model ‘Pubertal Status’ concepts from the Excel version of the CCPDD as shown in Figure 5 below. The data items ‘Male Genitalia Stage’, ‘Male Pubic Hair Stage’, ‘Female Breast Stage’, ‘Female Pubic Hair Stage’, ‘Date of Menarche’, ‘Age of Menarche’, and ‘Testicular Volume’ all existed in the Excel version of the CCPDD. Most of these data items were included as part of the Tanner Scales. This step could have potentially been simplified if the original model had been created in a more structured way and demonstrated one of the advantages of using a modelling tool over Microsoft Excel.

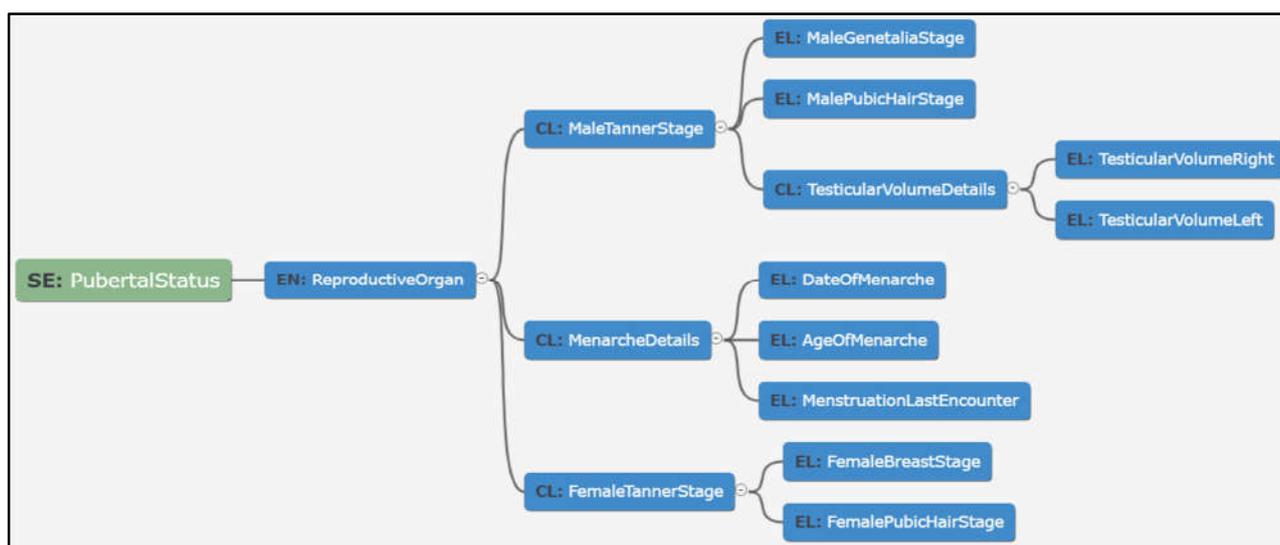


Figure 5. Example of modelling pubertal status concepts.

In the period between the finalisation of CCPDDv1 and modelling the pubertal status concepts described above, the CDISC published examples of how these concepts are represented in CDISC standards [24]. The pubertal status models were mapped to the new user guide within Direct, as shown in Figure 6 below. As with the Microsoft Excel version of the CCPDD, mapping to standards within the modelling tool is done manually by searching for the online link and copying and pasting it into the space provided. While this does not initially seem to offer an obvious advantage over Excel, it does provide a clear list of the standards associated with the model.

Mapping Type	Mapping URL	Mapping Description
CDISC	https://www.cdisc.org/system/files/members/standard/ta/TAUG-T1D_Pediatrics_and_Devices_v2.0.pdf	Tanner Staging is modelled in the CDISC Therapeutic Area Data Standards for Type 1 Diabetes - Pediatrics and Devices Modules

Figure 6. Image showing the CDISC mapping uniform resource locator (URL) and description. CDISC—Clinical Data Interchange Standards Consortium.

In summary, using a modelling tool facilitated the discipline of providing unambiguous definitions of data dictionary items which can be consumed and implemented downstream. The rigour of modelling helped not just the producers of the data dictionary, but the consumers of the data dictionary. The pilot has also facilitated the unexpected feedback loop into the standards body to help improve definitions, highlighting that standards bodies need to keep abreast of improvements in how these data dictionary items are now being modelled.

3.2. Data Visualisation

Unlike Microsoft Excel, the HTML version as shown in Figure 7 above, provided a summary of the modelled data dictionary in the clinical modelling tool aimed at providing a factsheet of the data content and how it was modelled. Additionally, the HTML version of data items enabled the clinician to have a clear understanding of the description of each data item, how they were represented, and the importance of each representation.

The screenshot displays the UNEW C4C interface. On the left, a navigation tree under 'CCPDD-Pilot 2 Models' shows a hierarchy: EN BODYMEASUREMENT (selected), CL BodySurfaceAreaDetails, EL BodySurfaceArea, CL BloodPressureMeasurement, EL SystolicBloodPressure, EL DiastolicBloodPressure, EL Weight, EL BMI, EL Height, and EL TotalBodyLength. The main panel shows the 'Model is in read only mode' and 'HTML' view. The entry details for 'BodyMeasurement' are as follows:

Entry Name : BodyMeasurement Version Status : VER00(Draft) Version Number : 0.1 [read less...](#)

Entry Description : A range of body measurements that may be recorded for a subject. This includes but is not limited to the measurement of height, weight, and head circumference of the subject. It also includes the manual or calculated measurement of the BMI and body surface area of the subject.

archetype_id : ISO-EN13606-ENTRY.BodyMeasurement.v0.1

Cardinality : 0..* Conformance : Required Meaning Binding : Domain :

Mapping Type	Mapping URL	Mapping Description
CDASH	https://www.cdisc.org/system/files/members/standard/foundational/CDASHIG%20v2.1-Final_Rev.pdf	Section 8.3.18
SDTM	https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-3/html#Vital+Signs	
FHIR Mapping	http://hl7.org/fhir/StructureDefinition/bodyweight	Body weight
FHIR Mapping	http://hl7.org/fhir/StructureDefinition/bodyheight	Body height
FHIR Mapping	http://hl7.org/fhir/StructureDefinition/bodylength	Body length
FHIR Mapping	http://hl7.org/fhir/StructureDefinition/bmi	BMI

Attestation Information : [read more...](#)

Audit Information : [read more...](#)

Other Attribute : [read more...](#)

Figure 7. Image of the HTML version of the modelled data dictionary. SDTM—Study Data Tabulation Model; FHIR—Fast Healthcare Interoperability Resources; CDASH—Clinical Data Acquisition Standards Harmonisation; HTML—hypertext markup language.

Figure 8 depicts the mind map of a modelled data dictionary item with the specific 'Entry', linked 'Elements' and 'Clusters' and the 'Clusters' linked to other 'Elements'. This mind map offered clearer visibility for the clinician on the data dictionary content being modelled compared to the Excel format. In contrast to the Excel worksheet, the mind map provided a schematic flow chart-like template with a clear guide to the clinician or clinical trial personnel on how data items were modelled and how they relate to each other.

3.3. Export of Data in ODM Format

Direcht allows the export of modelled data dictionary items through the ODM format into REDCap, as shown in Figure 9 below, and as such promotes interoperability of data through ensuring consistent eCRF design between trial sites. However, after the export of the modelled data dictionary items in ODM format into REDCap, some parts of the data dictionary items were found to be missing due to the inability of the format to allow parts of the modelled item that did not conform to CDISC standards to be exported. Such parts included the conformance and the cardinality of the modelled dictionary items. Hence, some data dictionary items which cannot be standardised in the CDISC format cannot be exported to REDCap. Furthermore, none of the modelled data dictionary items were able to be exported to Castor—another eCRF implementation—using the ODM format, as Castor at the time did not support the ODM. However, c4c has made significant progress through advocacy and collaboration with partners to influence the launch of a new update of Castor that will enable the exportation of the ODM format of modelled data dictionary

items. There was also an issue with spaces in field names when the models were exported to REDCap. Direct fields do not allow spaces (for, e.g., DateOfBirth), which means there were no spaces in these fields on the CRF. This was because element names in Direct become variable names in the back end of the tool. It would be beneficial to have more than one EDC to test the ODM import, as this would help to determine what was tool specific and what needs to be changed at the ODM or data modelling layer. There was also a limitation on the transfer of code lists in the ODM format, and they had to be put into the free text node.

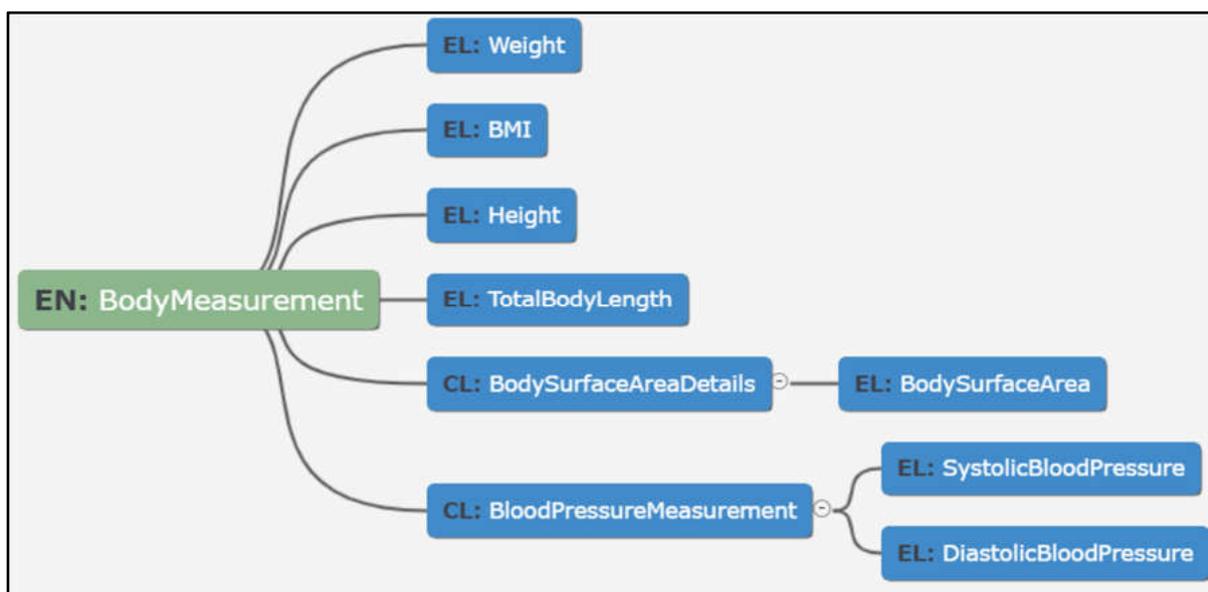


Figure 8. Image of the mind map of a modelled data dictionary. EN—entry; CL—cluster; EL—element.

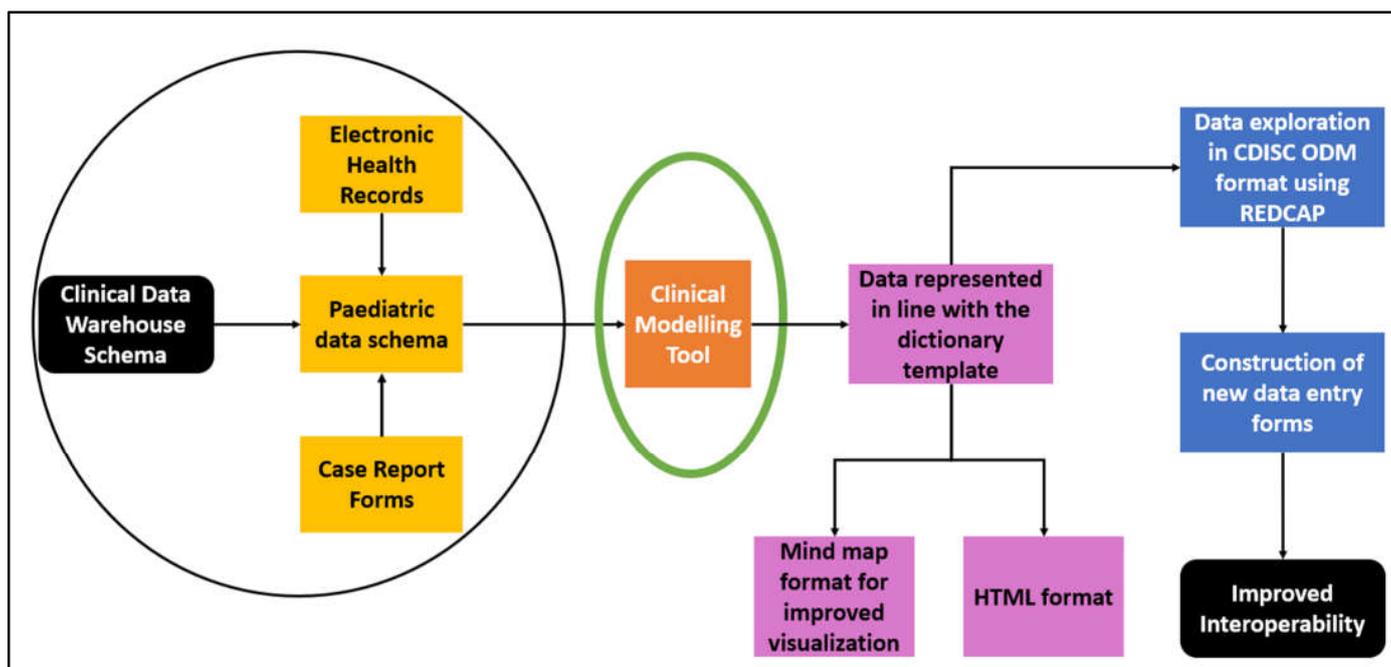


Figure 9. Schematic diagram of the role of a modelling tool in a clinical trial environment.

The team modelled the concept ‘PregnancyRelatedProcedures’ within Direcht and imported the ODM XML file into REDCap. REDCap was chosen as it is one of the more commonly used eCRF tools across the c4c consortium sites and claims to conform to the CDISC ODM standard [25]. It was therefore regarded as a suitable choice of eCRF to validate the Direcht ODM export. Overall, this export validation was successful; however, certain issues were identified:

- The structure of the data model created two CRFs within REDCap, rather than one;
- REDCap took the first variable in the XML file as the Subject ID; however, the model did not have a Subject ID, and as a result, the first field in the data model was not usable in the REDCap project;
- Data were entered into Direcht without spaces, for example, ‘DateOfBirth’. This format was exported into REDCap, affecting the user experience.

These results have prompted discussions between the c4c team, Direcht, and CDISC to determine at which point in the process they occur and how to address them. This work is ongoing and will influence the next stage of the pilot.

3.4. Findings from the First Pilot Study

The first pilot was launched in 2019 and was completed successfully in January 2020. The pilot validated the following:

- A robust, consistent, and systematic approach to modelling items in a non-standardised format can be achieved. This approach also helped in highlighting ambiguity in data definitions, which were otherwise difficult to identify.
- Using modelling tools provided the benefits of following a disciplined manner of structuring data with its attributes, cardinalities, associations, and values. This rigour in data representation benefited downstream tools such as EDC and reporting tools, providing consistent implementations.
- Modelling in a tool-based environment enabled viewing data in multiple dimensions such as tabular and graphical views to suit the user preference. This resulted in more engagement in reviews and discussions, which ultimately helped in richer and unambiguous data models, as shown in Figure 10 below:

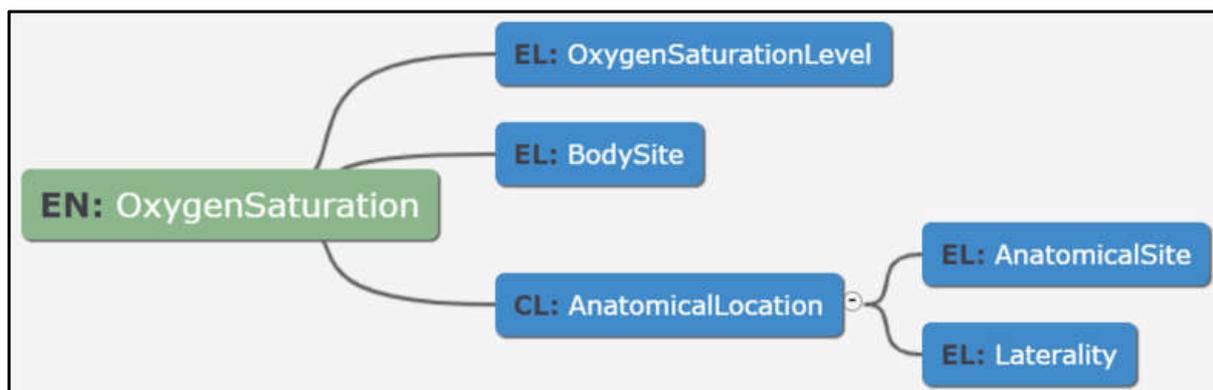


Figure 10. Modelled data dictionary item of oxygen saturation.

- Transformations to other standard formats including CDISC ODM can be achieved reducing the burden of manual mapping and human errors arising from such a process.
- Representing the data dictionary in the form of models and then exporting them to industry standards such as CDISC ODM enabled various EDC systems, e.g., REDCap, and used them to construct new data entry forms.

3.5. Benefits of Using a Clinical Modelling Approach over Microsoft Excel

Even though there were some challenges associated with using a modelling approach—as highlighted above—the following benefits were identified:

- Guided dictionary entry authors as to how to specify new clinical models (data dictionary entries);
- Acted as a collaboration platform for virtual team model creation, capturing design comments and decisions;
- Ensured the quality of each model specification so that all necessary modelling properties were specified and there was internal consistency;
- Enabled seamless linkage to external terminology systems for defining value lists, measurement units, etc.;
- Easily searchable and browsable library of models to avoid the creation of duplicates or overlapping models and to facilitate their use;
- Easy specialisation of models: use one as a base pattern to create a specialised version for a specified purpose, e.g., to adapt an adult model for children or a child model to neonates;
- Models were version managed, distinguishing draft versions and published versions;
- Models can be grouped into sets for convenient reuse and sharing;
- Exportable to CDISC ODM, for import into EDC systems, e.g., REDCap;
- Exportable to formats that can be used within EHR systems and RWD repositories;
- Held in a standards agnostic form, but conforming to ISO modelling standards to avoid vendor lock-in.

4. Discussion

The results from the pilots show the benefits of a data modelling tool over the widely used Microsoft Excel for representing data dictionary items. Benefits include:

- Enhanced user friendliness and overall usefulness due to visualisation tools. Mind maps enable clinicians and other clinical trial personnel to better understand the step-wise format by which data dictionary items are represented for use in clinical trials [26];
- The HTML summary provides a ‘fact sheet’ of the data dictionary items for enhanced clarity in line with the clinical registry template [27];
- A clearer way of representing a large list of data dictionary items;
- The flexibility to allow generic ‘Clusters’ and ‘Entries’ to be reused for other models.

In addition to facilitating and simplifying the CCPDD for clinicians, researchers, and data managers, Direcht brings important advantages for improving interoperability of data. The export of data models as ODM and XML files allows clinicians to effectively share data on a global level. Alignment with CDISC standards creates a widely accepted and officially recognised template for data dictionary items [28].

The XML export format would—in principle—permit c4c clinical models authored using Direcht to be imported to another ISO 13606 conformant archetype editor, avoiding the risk of vendor lock in. This will be an important consideration if the pilot is extended to a sustained clinical modelling approach for the rest of the project.

The successful export of data items to the CSIC ODM standard showed that Direcht ensured that every aspect of the modelled data item conformed with CDISC specifications. This is important because standardised data items boost intra usability and interoperability of data. However, this also posed a challenge for data items that could not be mapped to the ODM standard, meaning that Microsoft Excel was preferred in this context. Table 2 below summarises the implications of the clinical modelling tool pilot.

Table 2. Implications of the clinical modelling pilot.

Feature	Practical Implication
Interoperability between standards	Creates a bridge between the clinical and research worlds. Potential to combine standardised data from different sources.
User friendliness	Heightened understanding of the representation of clinical paediatric data items. Personnel less familiar with data standards can use the models, resulting in a wider pool of standardised data. This could be of particular importance to academic studies.
Automated eCRF creation potential	Researchers can export models into their EDC systems. This avoids any human error in applying standards and makes it easier for personnel, particularly those new to implementing data standards at the CRF level.

5. Conclusions

The pilots undertaken to date have confirmed that the use of a standardised but semantically agnostic clinical modelling tool offers multiple benefits in representing the c4c CCPDD items, compared with the more traditional Microsoft Excel-style resource. Tools such as *Direcht*, which can render the c4c CCPDD easier to use in clinical trials, carry significant added value for clinicians and researchers who are typically time-short and require assets which simplify processes and aid understanding. Beyond this, however, the tool offers major potential to address the current lack of standardisation—and, by extension, interoperability—of the data most often collected in a typical paediatric study. A greater volume of more easily pooled data should support more streamlining of future research, for instance in terms of adapted trial design—potentially reducing the use of placebo arms, for example. The potential to map the c4c CCPDD items to real world data, such as EHRs, is also powerful, and could facilitate activities such as extracting data from EHRs to populate eCRFs for a study, as well as using EHR or registry data to support post-marketing surveillance studies. The c4c consortium will continue to pilot use of this tool, to ascertain its added value for different stakeholders and, simultaneously, continue to build synergies between standards development organisations, stimulating adaptations which should benefit the wider paediatric health and research data communities.

Author Contributions: Conceptualisation, V.S., C.A., R.L., A.P., V.H., R.Q.S. and D.K.; data curation, C.A., R.L., A.P., R.Q.S. and D.K.; formal analysis, C.A., R.L., A.P., A.S., R.Q.S. and D.K.; funding acquisition, R.L.; investigation, C.A., R.L., A.P., R.Q.S. and D.K.; methodology, V.S., C.A., R.L., A.P., R.Q.S. and D.K.; project administration, R.L. and V.H.; resources, C.A., R.L., A.P., R.Q.S. and D.K.; software, R.Q.S.; supervision, R.L. and V.H.; validation, C.A., R.L., A.P., A.S., R.Q.S. and D.K.; visualisation, V.S., C.A., R.L., A.P., V.H., A.S., R.Q.S. and D.K.; writing—original draft, C.A., R.L., A.P., V.H., R.Q.S. and D.K.; writing—review and editing, V.S., C.A., R.L., A.P., V.H., A.S., R.Q.S. and D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 777389. The Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA.

Data Availability Statement: Data sharing not applicable. No new data were created or analysed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: Rahil Qamar Siddiqui, co-founder of Sidqam Ltd., which has developed the *Direcht* Modelling Tool used as part of the modelling work performed by the *conect4children* (c4c) Cross Cutting Paediatric Data Dictionary (CCPDD) team.

Disclaimer: The publication reflects the authors' view and neither IMI nor the European Union, EFPIA, or any associated partners are responsible for any use that may be made of the information contained therein.

References

1. Harris, P.A.; Taylor, R.; Thielke, R.; Payne, J.; Gonzalez, N.; Conde, J.G. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **2009**, *42*, 377–381. [[CrossRef](#)] [[PubMed](#)]
2. Turner, M.A.; Hildebrand, H.; Fernandes, R.M.; De Wildt, S.N.; Mahler, F.; Hankard, R.; Leary, R.; Bonifazi, F.; Nobels, P.; Cheng, K.; et al. The conect4children (c4c) Consortium: Potential for Improving European Clinical Research into Medicines for Children. *Pharmaceut. Med.* **2021**, *35*, 71–79. [[CrossRef](#)] [[PubMed](#)]
3. Conect4children (c4c). Available online: <https://conect4children.org/> (accessed on 20 January 2022).
4. CDISC. Therapeutic Area Data Standards for Type 1 Diabetes—Pediatrics and Device Modules. Available online: <https://www.cdisc.org/system/files/members/standard/ta/TAUG-> (accessed on 20 January 2022).
5. Health Level Seven International. HL7 Standards. Available online: <https://www.hl7.org/> (accessed on 20 January 2022).
6. Ohmann, C.; Banzi, R.; Canham, S.; Battaglia, S.; Matei, M.; Ariyo, C.; Becnel, L.; Bierer, B.; Bowers, S.; Clivio, L.; et al. Sharing and reuse of individual participant data from clinical trials: Principles and recommendations. *BMJ Open* **2017**, *7*, e018647. [[CrossRef](#)] [[PubMed](#)]
7. Project, T.Y. Yale University Open Data Access (YODA) Project Procedures to Guide External Investigator Access to Clinical Trial Data. Available online: <https://yoda.yale.edu> (accessed on 20 January 2022).
8. De Moor, G.; Sundgren, M.; Kalra, D.; Schmidt, A.; Dugas, M.; Claerhout, B.; Karakoyun, T.; Ohmann, C.; Lastic, P.Y.; Ammour, N.; et al. Using electronic health records for clinical research: The case of the EHR4CR project. *J. Biomed. Inform.* **2015**, *53*, 162–173. [[CrossRef](#)] [[PubMed](#)]
9. Claerhout, B.; Kalra, D.; Mueller, C.; Singh, G.; Ammour, N.; Meloni, L.; Blomster, J.; Hopley, M.; Kafatos, G.; Garvey, A.; et al. Federated electronic health records research technology to support clinical trial protocol optimization: Evidence from EHR4CR and the InSite platform. *J. Biomed. Inform.* **2019**, *90*, 103090. [[CrossRef](#)] [[PubMed](#)]
10. Griffon, N.; Pereira, H.; Djadi-Prat, J.; Garcia, M.T.; Testoni, S.; Cariou, M.; Hilbey, J.; N'Dja, A.; Navarro, G.; Gentili, N.; et al. Performances of a Solution to Semi-Automatically Fill eCRF with Data from the Electronic Health Record: Protocol for a Prospective Individual Participant Data Meta-Analysis. *Stud. Health Technol. Inform.* **2020**, *270*, 367–371. [[CrossRef](#)] [[PubMed](#)]
11. European Commission, Directorate-General for the Information Society and Media. *Semantic Interoperability for Better Health and Safer Healthcare: Deployment and Research Roadmap for Europe*; Virtanen, M., Ustun, B., Rodrigues, J., Stroetmann, V., Surjan, G., Rector, A., Stroetmann, K., Lewalle, P., Zanstra, P.E., Kalra, D., Eds.; European Commission: Brussels, Belgium, 2013.
12. Beale, T. Archetypes: Constraint-based Domain Models for Future-proof Information Systems. 2002. Available online: https://www.researchgate.net/publication/237033734_Archetypes_Constraint-based_Domain_Models_for_Future-proof_Information_Systems (accessed on 20 January 2022).
13. Garde, S.; Chen, R.; Leslie, H.; Beale, T.; McNicoll, I.; Heard, S. Archetype-Based Knowledge Management for Semantic Interoperability of Electronic Health Records. *Stud. Health Technol. Inform.* **2009**, *150*, 1007–1011. [[CrossRef](#)]
14. Moreno-Conde, A.; Moner, D.; Cruz, W.D.; Santos, M.R.; Maldonado, J.A.; Robles, M.; Kalra, D. Clinical information modeling processes for semantic interoperability of electronic health records: Systematic review and inductive analysis. *J. Am. Med. Inform. Assoc.* **2015**, *22*, 925–934. [[CrossRef](#)] [[PubMed](#)]
15. Kalra, D.; Tapuria, A.; Austin, T.; De Moor, G. Quality requirements for EHR Archetypes. *Qual. Life Through Qual. Inf.* **2012**, *180*, 48–52. [[CrossRef](#)]
16. Ahn, S.; Huff, S.M.; Kim, Y.; Kalra, D. Quality metrics for detailed clinical models. *Int. J. Med. Inform.* **2013**, *82*, 408–417. [[CrossRef](#)] [[PubMed](#)]
17. ISO. ISO 13606-2:2019. *Health Informatics—Electronic Health Record Communication—Part 2: Archetype interchange specification*. Available online: <https://www.iso.org/standard/62305.html> (accessed on 20 January 2022).
18. ISO. ISO 13606-1:2019. *Health Informatics—Electronic Health Record Communication—Part 1: Reference Model*. Available online: <https://www.iso.org/standard/67868.html> (accessed on 20 January 2022).
19. Moreno-Conde, A.; Austin, T.; Moreno-Conde, J.; Parra-Calderon, C.L.; Kalra, D. Evaluation of clinical information modeling tools. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 1127–1135. [[CrossRef](#)] [[PubMed](#)]
20. Tapuria, A.; Bruland, P.; Delaney, B.; Kalra, D.; Curcin, V. Comparison and transformation between CDISC ODM and EN13606 EHR standards in connecting EHR data with clinical trial research data. *Digit Health* **2018**, *4*, 2055207618777676. [[CrossRef](#)] [[PubMed](#)]
21. Direct. Available online: <https://www.direct.com/> (accessed on 20 January 2022).
22. Martinez-Costa, C.; Menarguez-Tortosa, M.; Fernandez-Breis, J.T. Towards ISO 13606 and openEHR archetype-based semantic interoperability. *Stud. Health Technol. Inform.* **2009**, *150*, 260–264. [[PubMed](#)]
23. Austin, T.; Sun, S.; Hassan, T.; Kalra, D. Evaluation of ISO EN 13606 as a result of its implementation in XML. *Health Inform. J.* **2013**, *19*, 264–280. [[CrossRef](#)] [[PubMed](#)]

24. CDISC. Therapeutic Area User Guides. Available online: <https://www.cdisc.org/standards/therapeutic-areas> (accessed on 20 January 2022).
25. Yamamoto, K.; Ota, K.; Akiya, I.; Shintani, A. A pragmatic method for transforming clinical research data from the research electronic data capture “REDCap” to Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM): Development and evaluation of REDCap2SDTM. *J. Biomed. Inform.* **2017**, *70*, 65–76. [[CrossRef](#)] [[PubMed](#)]
26. Gregson, C. Clinical trial data visualisation. *Trials* **2015**, *16*, P187. [[CrossRef](#)]
27. Hassanzadeh, O.; Lim, L.; Kementsietsidis, A.; Miller, R.J.; Wang, M. LinkedCT: A Linked Data Space for Clinical Trials. *arXiv* **2009**, arXiv:0908.0567.
28. Meineke, F.A.; Staubert, S.; Lobe, M.; Winter, A. A comprehensive clinical research database based on CDISC ODM and i2b2. *Stud. Health Technol. Inform.* **2014**, *205*, 1115–1119. [[PubMed](#)]