The Model of Reading

Modelling principles, Definitions, Schema, Alignments

Version	2.0
Last Update	17/07/2019
Editor	Alessio Antonini – The Open University, Milton Keynes, UK - alessio.antonini@open.ac.uk
Contributors	François Vignale - Le Mans Université, Le Mans, France francois.vignale@univ-lemans.fr
	Guillaume Gravier - IRISA, Rennes, France guig@irisa.fr
	Brigitte Ouvry Vial - Le Mans Université, Le Mans, France Brigitte.ouvry-vial@univ-lemans.fr

Executive Summary

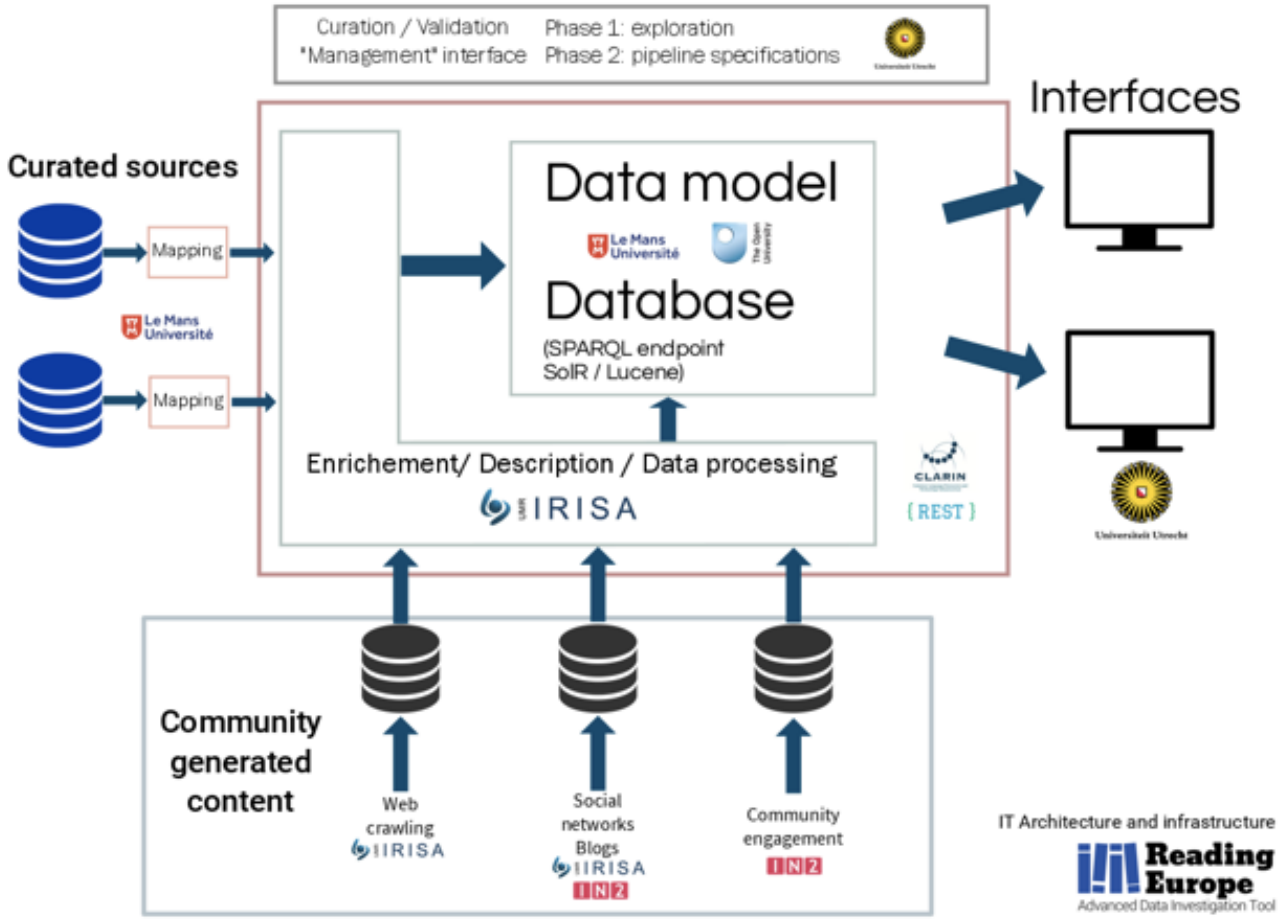
This technical report introduces the data model developed to address the systematic collection and use of reading experiences in READ-IT project. The model of reading presented in this document is meant to inform the development of the READ-IT database and tools. This document describes the methodological approach and design principles adopted in the development of the model of reading. Furthermore, this technical report describes the content of the first version of the data model of the reading experience, including a preliminary analysis of the alignments between READ-IT model of reading with CIDOC-CRM, FRBRoo, FoaF and Schema.org.

Index

Introduction ... **3**
The Model of Reading ... **6**
Reading: a Multifaceted Phenomenon ... 7
Action of Reading ... 8
The Agent of Reading and its States ... 10
Reading, Reading Session and Reading Experience ... 12
Experience of Reading ... 13
The Context of Reading ... 16
The External Context of Reading ... 16
The Personal Context of Reading ... 17
READ-IT Data Model - Reading Experience ... **18**
The Model of Reading - V1.6 ... 20
Classes of READ-IT data model ... 20
Relations of READ-IT data model ... 22
Alignment with Standards and Resources ... 23
Reader & Person ... 24
Medium & Content ... 25
States of Reading ... 28
Situation of Reading and CIDOC-CRM & Schema.org ... 28
Using the Data Model ... **29**
Annotations Metadata in READ-IT ... 29
W3C Web Annotation Data Model ... 29
Example 1 ... 30
Example 2 ... 33
Design Patterns ... 35
Reading to an audience ... 35
Non-readings, attempts to read and brief readings ... 36
Comparative readings ... 37
Cumulative outcomes of sequential readings ... 38
First and last impressions of reading ... 39
Conclusions and Future Developments ... **40**
Annexes ... **41**
Annex 1 - Question & Answers ... 41
Is the reader's writing habit within the scope of the model? ... 41
The reader is defining his/her experience by comparison. Does the model support comparison between experiences? ... 41
In case the evidences of reading are meant for someone, e.g. the author, how this information could be represented? ... 41
In the evidence, the reader is describing emotional aspects of the content and not of his/her personal reading experience. Is this information in the scope of READ-IT data model? ... 41
What is the minimum set of information required to have a reading experience entry? ... 41
How experimental data should be encoded in the model? ... 41
My evidence reports of reading social media, blog posts or other contents which are not books, how should that be encoded in the data model? ... 41
How should I specify if a reader is an expert and/or if the reader is reading as part of his/her professional activities? 41
Is there a distinction between fiction and non-fiction content? ... 42
How can I quantify the reader's engagement? ... 42
Annex 2 - References ... 43
Annex 3 – Changelog ... 45

Introduction

Developing a model of the reading experience is the first step toward the implementation of the READ-IT architecture. The model plays four main roles: 1) providing a common meta-language of reading experience, 2) guiding the development of READ-IT database, 3) supporting the interdisciplinary research addressing the macroscope questions [Hitchcock, T. 2014] about reading in Europe, and 4) supporting the multidisciplinary research of the individual research groups.



READ-IT general architecture.

There is as yet no unifying theory of data modelling in the Humanities, where research has focused on the modelling of cultural artefacts and not on the modelling of processes [Flanders and Jannidis 2016]. The READ-IT data model builds upon outcomes from previous DH research on the Cultural Heritage of experiences[1], with the aim of overcoming a number of crucial conceptual and technical limitations of these projects, such as their limited geographical range and their lack of integration of multimodal digitized sources of evidence (images, texts, audio sources, computer-mediated communications).

The READ-IT data model concentrates on a more fundamental limitation of previous attempts to model experiences. In particular, the UK RED and LED projects adopted a practice-driven approach with a focus on the management on the data sources. Their overall aim was to enable crowdsourcing and collaborative annotation, with an emphasis on representing the content (e.g. music performance, book fragment) [Adamou et al., 2018] rather than the personal perspective of the experience. Thus, these models focus on the process of managing

[1] Such as the UK-Reading Experience Database (2006-present), the Reading in Europe: Contemporary Issues in Historical and Comparative Perspectives project (2014-2017) and the Listening Experience Database project (2012-present).

research data rather than on the experience which was modelled as a reference connecting a piece of evidence in a source to a content, e.g. a text paragraph with a recording segment.

In this project, we consider reading as the set of physical and cognitive activities related to the direct engagement of a person (reader) with a meaningful linguistic expression (content) through physical or digital means (support).

In the design of READ-IT data model we adopt a pragmatic approach. Our aim is not to define a theory of reading but a conceptual framework which could define a common ground among the different research strands. Following, the READ-IT data model is specifically oriented to the implementation of READ-IT database as collector of both manual and automated annotations of reading experience, and enabler of the different research programs. We address this challenge through a transdisciplinary approach by attempting to interconnect elements from different theories and rebuilding these different points of views in a coherent and consistent data model. The modelling is informed by theories and analysis about reading, driven and tested using the sources collected so far by researchers in READ-IT consortium and previous projects, and validated through the research questions of Humanities and Digital Humanities researchers. Furthermore, we adopt best practices and stands about the scope of READ-IT projects and the data sources we aim to address. In general, we summarise the approach to modelling in READ-IT with the following principles:

Multi-facet approach. Addressing the structuring of evidence of reading conceived as action, experience and event. As action, reading is performed by an actor (reader) who intentionally interacts with a content through a support, with prior aims and subsequent outcomes. As experience, reading is the cognitive journey of the reader, beginning with their skill, social, aesthetic dispositions and complicated by interfering environmental factors. As event, reading takes place in a socio-cultural context in a place and time meaningful for the reader.

Database of evidences. Providing the means to annotate a source (textual, visual, or other) in order to collect evidence that supports the retrieval and study of specific aspects of reading. It does not provide a specific analytical framework to interpret sources nor to manage distinctions between reality and evidence in the text.

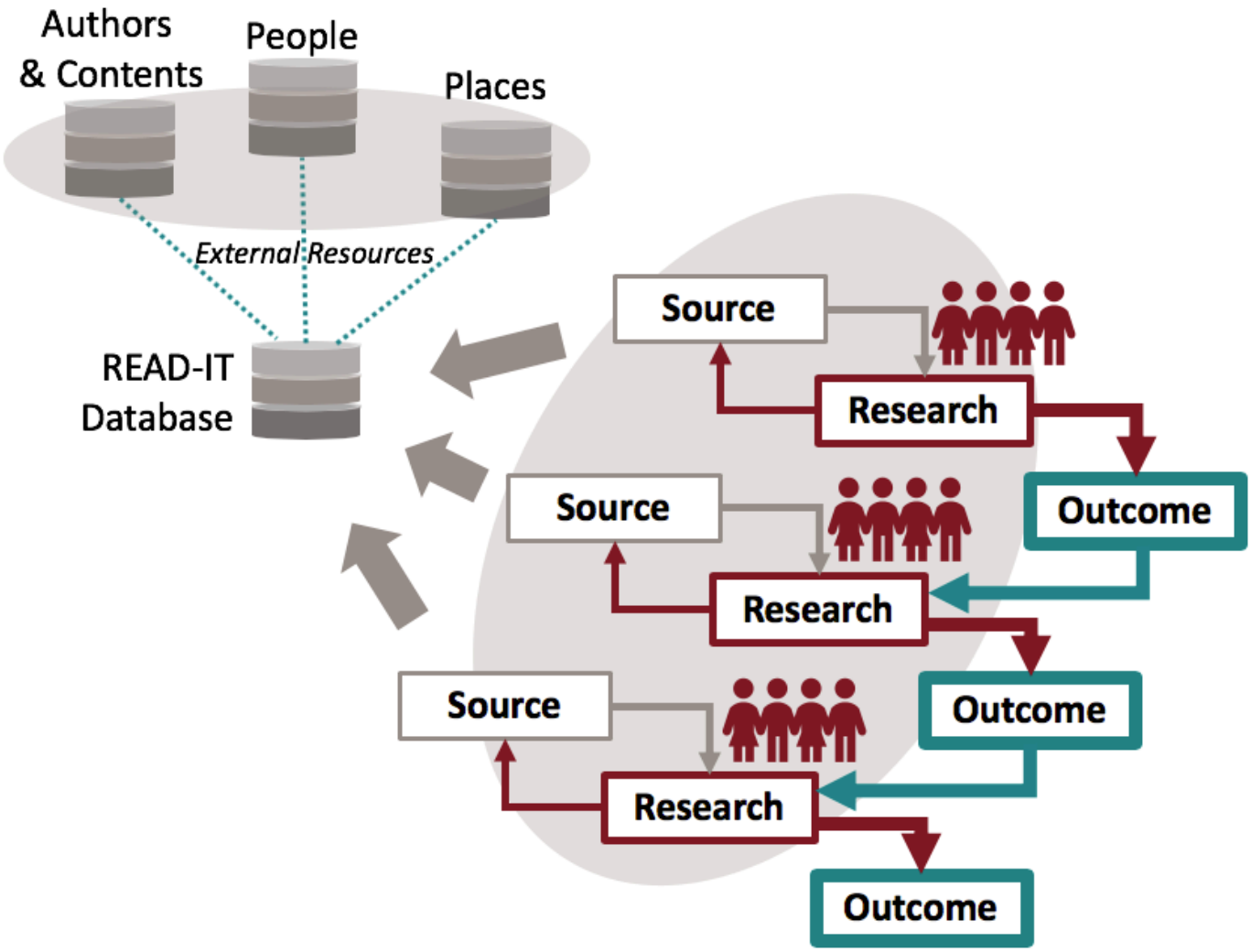
Data-driven and use-case driven. Data-driven validation i.e. based on the analysis of evidences found in collected corpora. The testing of the model is use-case driven. Through a dialogue with the DH researchers within READ-IT, we stress the ability of the model to represent the relevant aspects of sources and to address diverse research questions.

Standards and external resources. Use of existing standard and alignment with existing resources. Where concepts are addressed by established standards, such as “Person”, “Content” or “Event”, the model adopts and extends such standards, enabling seamless linking and reuse of these external resources.

Supporting computational approaches and manual annotations. The model considers data features and research questions way beyond the current state of the art of computational techniques of data analysis. With a focus on the complexity of reading as multi-facet phenomenon, the aim of the model is to provide support as the techniques evolve, and to the most sophisticated manually-annotated corpora which will contribute to such evolution.

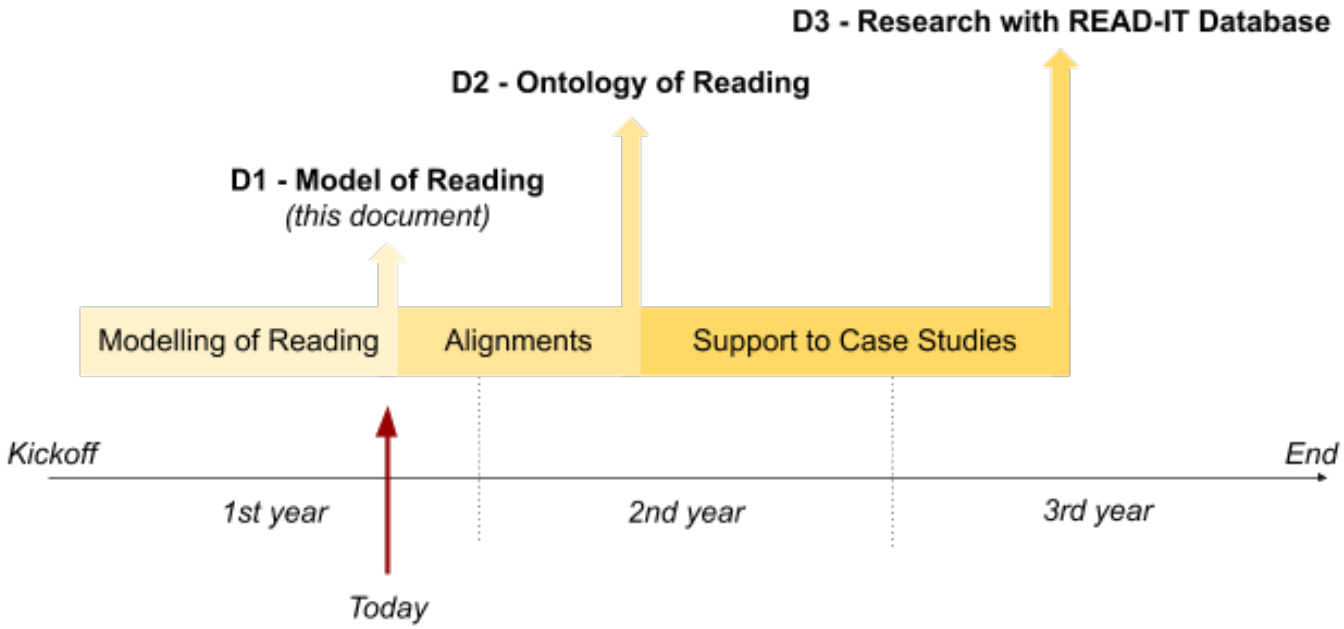
Multimodal multi-languages digitized sources. Support sources of different languages and of different types, e.g. diaries, tweets, images, recordings.

It is worth to highlight that one of the objectives of READ-IT is to facilitate the re-use of the annotation of sources and of the outcomes of case studies in the formulation of new case studies and research programs. In this frame, the development of READ-IT data model must be oriented primarily to the generation of new research data (nor organisation of data) based on a common understanding. Indeed, the role of technology will be facilitating the adoption of a common set of terms, concepts and practices (based on the use of READ-IT toolkit) among different research groups and disciplines with a focus on reading.



READ-IT database is meant to include the sources and annotation of produced by READ-IT case studies, internal and external partners generated through READ-IT technological ecosystem. The objective is to achieve an interoperability between sources, outside the scope of specific case studies.

This document reports the first version of model of reading developed through the study of a selection of sources and by engaging the researchers involved leading the READ-IT case studies. Specifically, the next section reports the approach used in the modelling activity and describes the concepts and relations of the mode of reading. The following section addresses the relations between READ-IT model of reading and relevant knowledge structures and schema. The section Using the Data Model addresses the use of the model for the annotation of sources and provides examples and design patterns emerging from the analysis of sources. Lastly, we provide concluding remarks about the scope of the model and the future developments.



This document is the first of three Deliverables (D1 - Model of Reading) and a project milestone, reporting the activities of the first year of READ-IT related to the modelling of reading experience. The following technical report will address the formal encoding of an ontology of reading, D2 - Ontology of Reading. The last one, D3- Research with READ-IT Database, will report design patterns, issues and methodologies emerging from the use of READ-IT data in the development of READ-IT case studies.

The Model of Reading

The model of reading will be used to inform the development of READ-IT database and annotation tools. The annotation tool and database will be used to generate data (annotation of sources) about reading experience. The annotation of sources will be used to develop READ-IT case studies and to address the interdisciplinary questions of READ-IT.

In this frame, the requirements for the development of the model of reading are related to three main capabilities:

1. representing (annotate) the different factors of reading experience emerging from the study of sources
2. integrating of the data generated from the different types of sources (e.g. diaries, correspondences, pictures)
3. retrieving the data required for addressing the research questions of the case studies

In the development of the model of reading, we followed an approach driven by sources, informed by theory and validated through case studies (for more information about the approach see http://oro.open.ac.uk/60762/). Driven by sources, as most of the concepts of of the model emerged from the study of sources about the reading experience. Informed by theory, we use theories about reading to guide the integration of the emerging concepts and formulate hypothesis to fill the potential gaps in the model. Lastly, validated through case studies, we consider the adoption and effectiveness of the model in the case studies as the criteria of success of the modelling activity.

At this stage of the project (end of the first year of three) we performed two cycles of 1) study of the sources 2) modelling and 3) engagement of researchers. It is worth considering that the annotation tools, database and case studies are currently under development, thus the validation stage had been done on the general definition of the case studies through their discussion and annotation sessions. The model we present in this document is a proposal, that will be further refined during the next two years of the project.

Reading: a Multifaceted Phenomenon

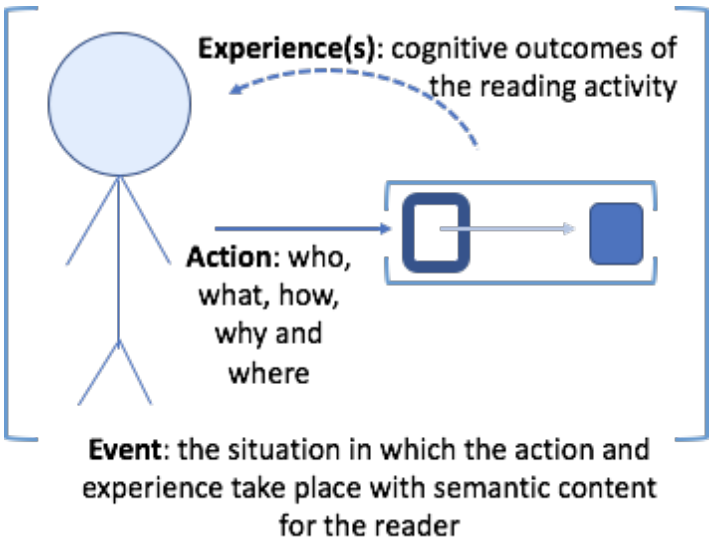
In the modelling of reading, the analysis of sources about reading experiences guided the definition of concepts. The concepts emerging from the analysis of sources are depending on the specific type of source. To overcome the fragmentation and provide a complete and consistent model, we had to integrate the analysis of sources with a conceptual analysis of the phenomenon of reading.

The analysis of reading had been informed by existing works on the field, and by defining three axes of analysis. Specifically, we adopted a reader-centred perspective considering reading as the action of interacting with a content through a support, as an experience resulting from the action, and as an event among the events of the reader's life. We adopted these three perspectives as lenses to analyse reading:

- As an **Action**, reading requires an actor (the Reader), an object (the Content) and the means of the action (the material Medium). Furthermore, reading as action does have a background (premises) related to the activities and aims of the reader and outcomes of the action regarding outputs, achievements impacting on the reader's activities.

- As an **Experience**, reading involves a set of cognitive changes resulting from the action of reading. Furthermore, the experience is also the result of the sum of factors interfering and altering the relations between reader, content, support and therefore the cognitive outcomes of reading.

- As an **Event**, reading is a meaningful occurrence in the reader's life, intentionally caused by the reader. As part of the reader's life, reading is affected by previous and co-occurring events and affecting the following events.



The action of reading is the trigger of the experience of reading. Action and experiences are embodied within a context, events, in which the reading is both object of and source of interference.

The adoption of these three axes resulted in the definition of three macro structures:

1. The action of reading, actor, object and means (tools), and its implementation process

2. The experience as changes in the reader as result of the reading activity

3. The event as context and as landmark in the reader's life

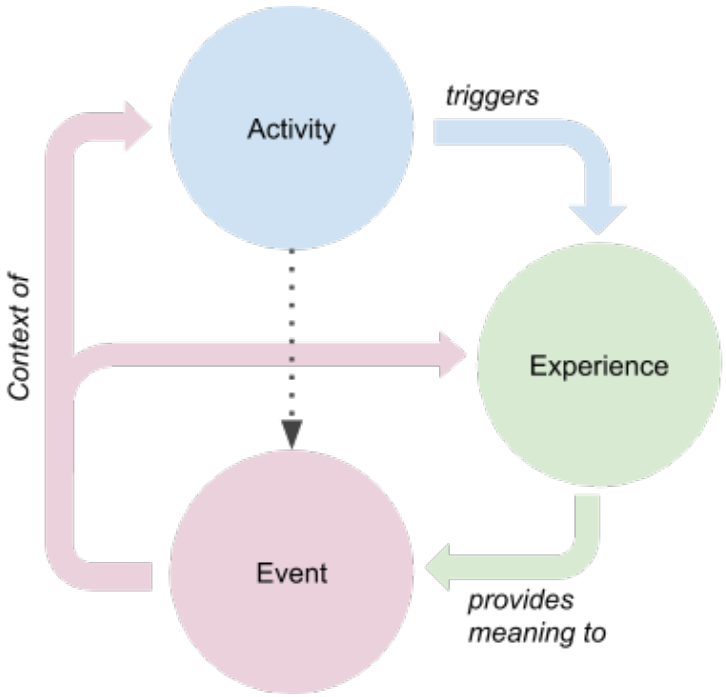
Following, these three macro structures were interconnected by adopting a phenomenological perspective of reading. Specifically, we consider:

a. the activity as the trigger of experience,

b. the experience as the development of meaning (e.g. semantic, emotion) which characterises the reading as event

c. events (reading and not reading) as context in which activities are performed

This high-level structure had been then used to define the relations between the macro-structures of the model of reading.



Connections between activity, experience and event.

Lastly, we had to take into account a set of assumptions and results from previous studies on reading, such as a listening a reading aloud is not a reading experience and that a reader is a human being.

Action of Reading

Reading is an action aimed to achieve the effect of reading a content. As action, reading implies the existence of an agent Reader with the objective of accessing a Content, with the support of a Medium. The action of reading is performed through a process (Reading Process) in which the Reader is *engaged in* and involving Medium and Content (a Reading Resource).

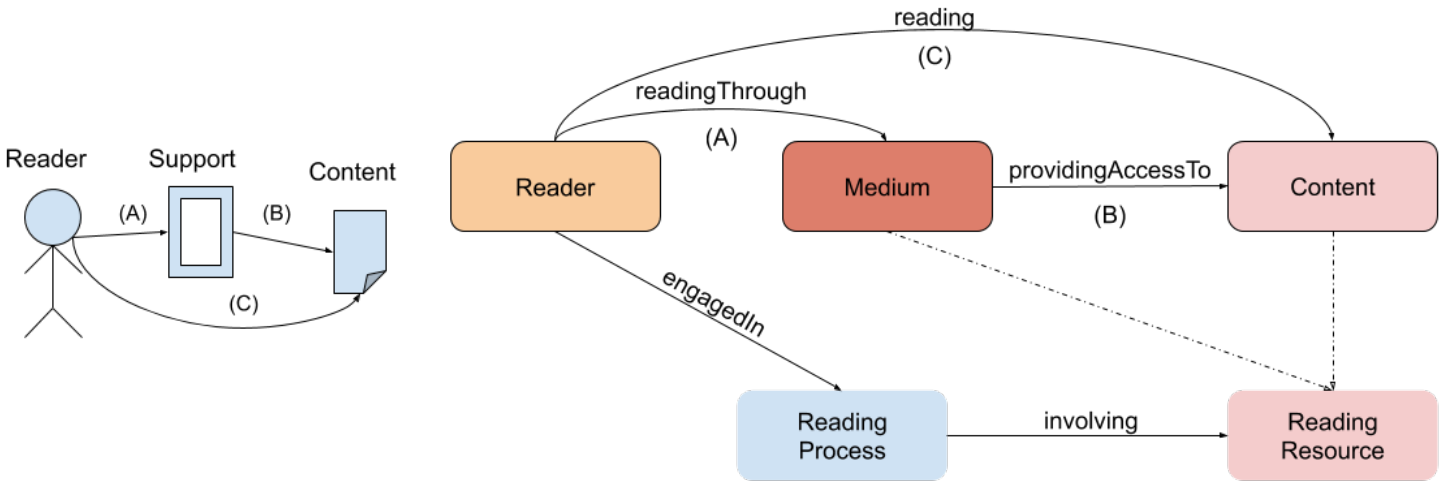
The act of reading encompasses three distinct types of interactions between Reader, Medium and Content:

(A) Reader *reading through* Support

(B) Support providing access to Content

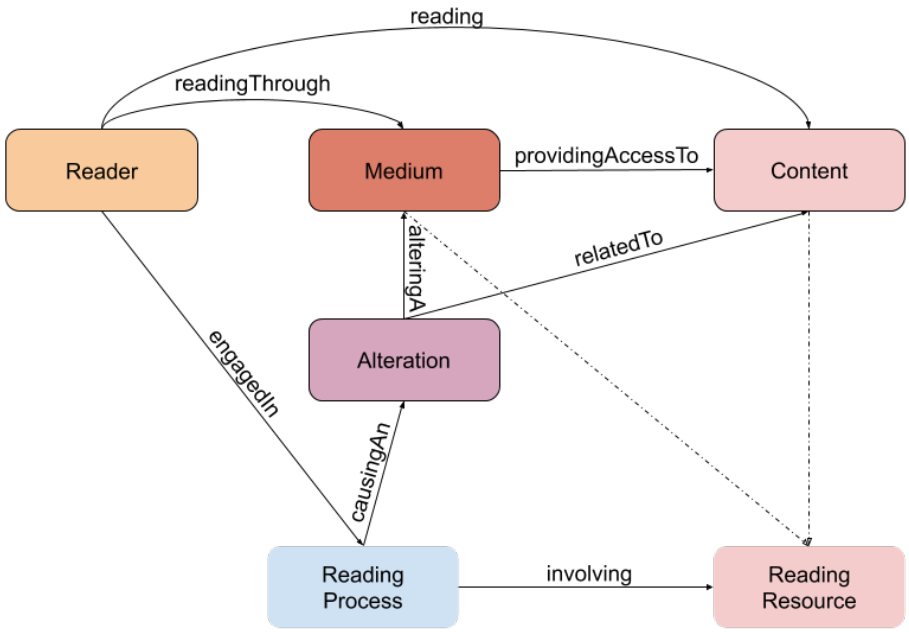
(C) Reader *reading* Content

We do consider the action of reading as implemented through a Reading Process, in which the Reader is *engaged with, involving* a Reading Resource.



On the left, the three-fold interaction of reading: (A) Reader with Support, (B) Reader accessing to the Content through the Support, (C) Reader accessing to the Content. To the Right, the schema encoding the three interactions.

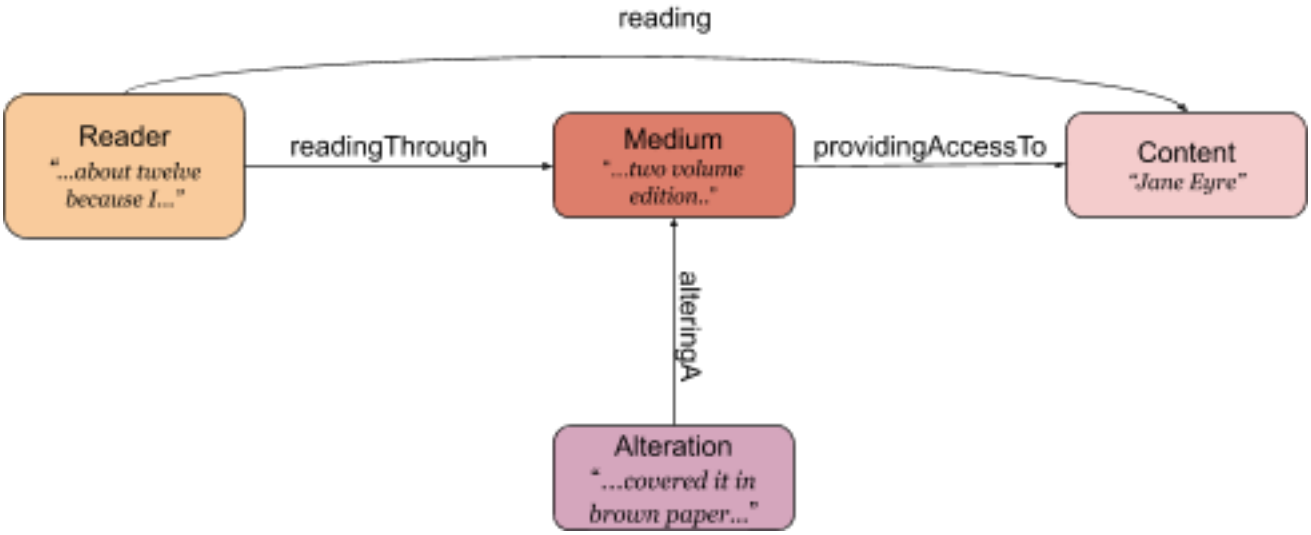
One of the possible outputs of a Reading Process is the Alteration of a Support. For instance, a Reader inscribes or attach a "note" (Alteration) *altering a* Support in *related to* a Content.



The schema encoding Reader, Medium and Content and Alteration (generalised as Reading Resource), and their relations with a Reading Process.

Example of reading experience including information about the reader, content, support and alteration:

"... it would have been about twelve because I read Jane Eyre ... and I remember it was my mother's two volume edition and we'd covered it in brown paper ..." Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.



During the analysis of sources, we identified testimonies about the "*engagement*" of the reader in the reading process. Specifically, so far, we identified expressions of a general engagement in reading and a type of engagement strictly related to the content, "*transportation*".

Extract including information about the engagement of the reader:

"*... Oh yes, oh, I remember* Jane Eyre ... *I just was swallowed up by it* ..." Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.

Extract including information about the transportation of the reader

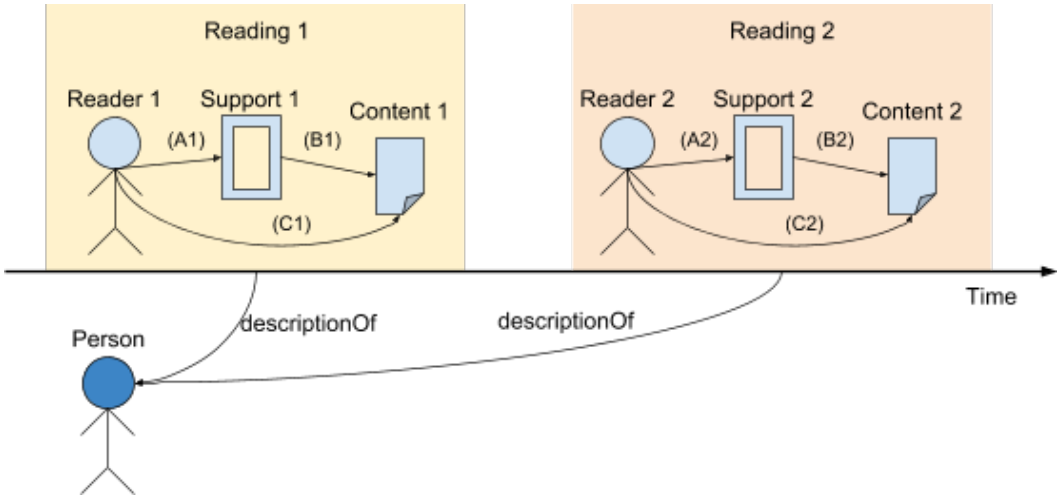
"... But did you identify with Jane Eyre? Yes, yes, to a certain extent, I think, but I liked her and I, I, I wanted to follow her on her journey and on her adventure ..." Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.

In the model of reading, we take into account *Engagement* and *Transportation* as properties of the Reading Process. The characterisation of *Engagement* and *Transportation* are still object of research.

The Agent of Reading and its States

An action is embedded in a specific context (time, place) and so is the agent. With Reader we indicate the state of the Person at the time of reading. The Person is the agent, while the Reader is one of the multiple states of the Person.

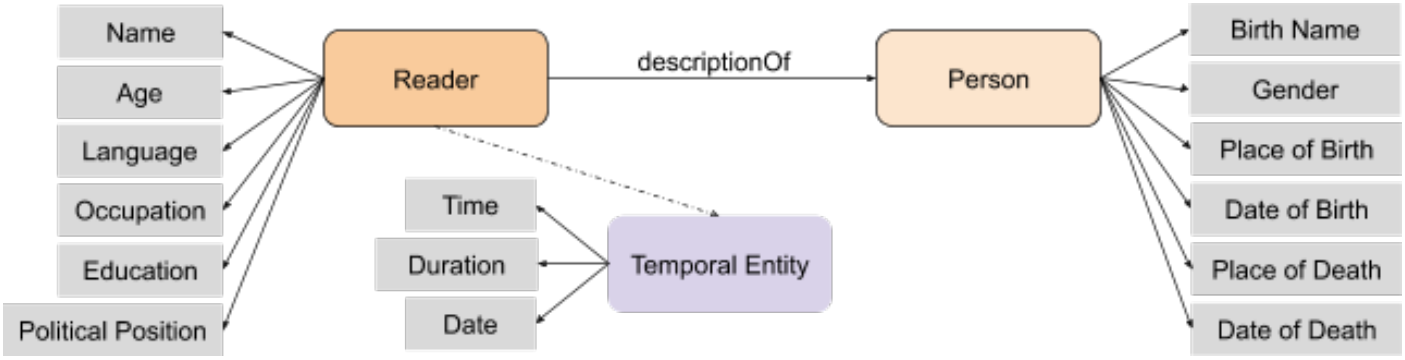
The rational is that the agent (Person) is acting under specific conditions and constraints, such as education level, beliefs, occupation. These conditions are of relevance for the study of the specific reading experience. Thus, we keep track of them with the concept of Reader. Furthermore, by keeping track of the different states of a Person, we can keep track of reading experience of the same individuals during time, see next figure.



A Person is the agent of reading acting under a specific state we represent with the concept of Reader. Thus, the interactions with Medium and Content happen under different conditions as the state of the Person changes (Reader1 to Reader2).

In this frame, Reader and Person are characterised by different features. A Person is characterised by "static" feature, not subject to change, while a Reader is characterised by the dynamic (variable) features of a human.

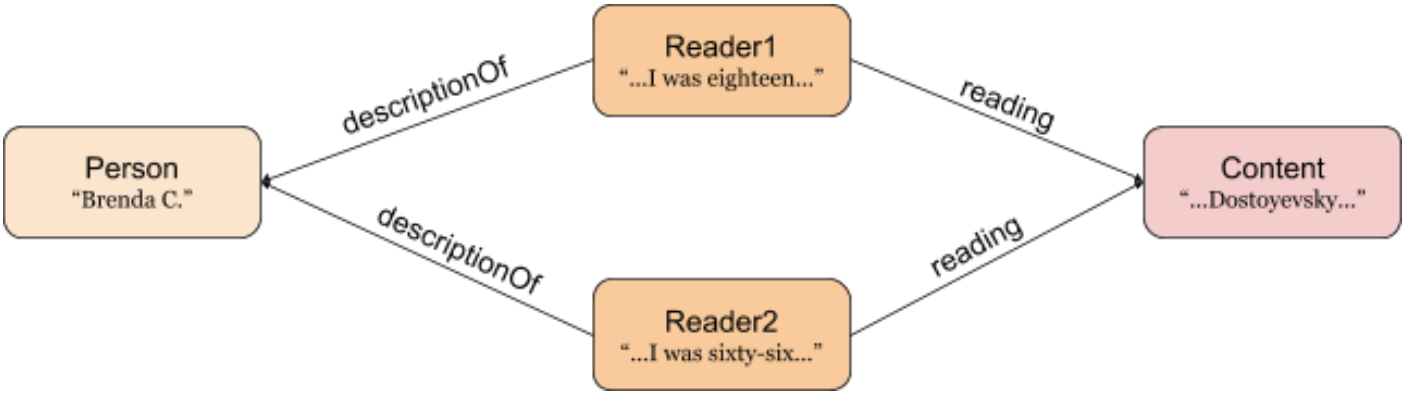
Person - Static Features	**Reader - Dynamic Features**
• Birth name • Gender • Date of birth • Place of birth • Date of death • Place of death • …	• Name • Age • Occupation • Education • Language • Residency • Political stand • Gender identity • Physical condition • …



The schema representing Reader and Person and Temporal Entity.

Example including information about the reading of the same content at different ages:

> *"I was reading the Dostoyevsky again, I read it in the sixth form, and this is another aside, but I was very taken with Raskolnikov, you see, when I was eighteen. I thought he was a bit of a, you know. I read it again when I was sixty-six…"* Extract from Memories of fiction, Brenda C. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/brenda-part-2.



During the analysis of sources, we identified testimonies about the reading *habits* of the reader. In our understanding of the testimonies, a reading habit is the intentional description of a class of reading experiences. The reader describes a habit providing a description of the features in common between a set of reading, such as location, time, genre, judgements about the content or the interaction with the medium.

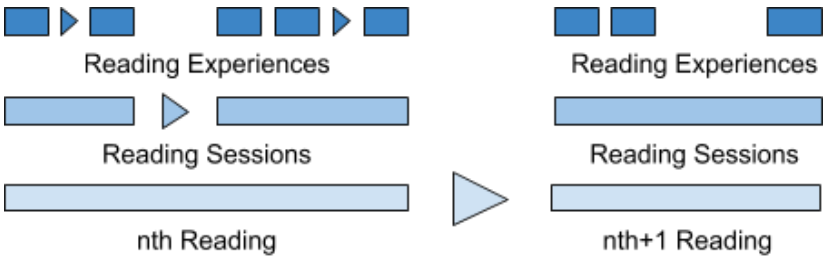
Extract including information about a *Habit* of the reader

> *"... I just was bored with the sort of characters, really, just ... Did you finish it? No, which is ... So, if you don't like something, you don't finish it? It's pretty rare that I don't finish something ..."* Source : Memories of fiction, Johanna W. interview part 1 [transcribed], 2015, https://soundcloud.com/memoriesoffiction/johanna-williams-1.

In the model of reading, we take into account the *Habit* as a property of the Reader. The characterisation of *Habit* is still object of research.

Reading, Reading Session and Reading Experience

Reading can have a long duration and therefore be performed in multiple reading sessions (Session). Furthermore, during each session the reader can have multiple reading experiences (Experience) in relation to the unravelling of the content, e.g. story twists, new characters or arguments.



A conceptual schema representing the dynamic of readings, relating Readings, Sessions and Experiences in terms of sequence and of inclusions.

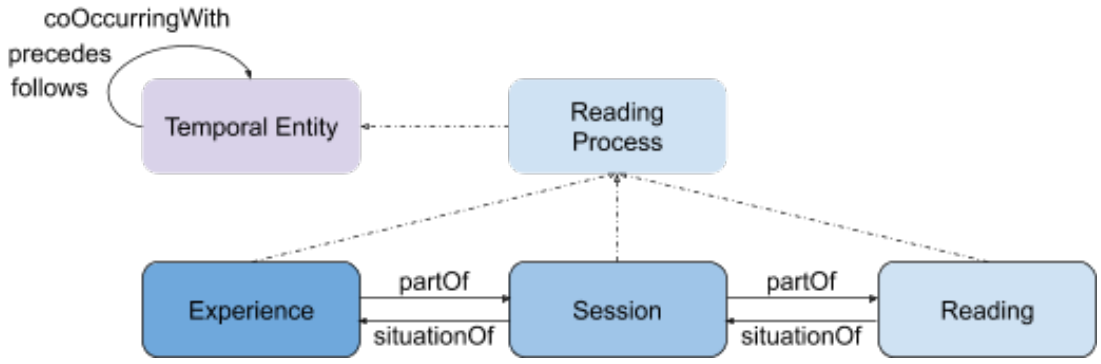
In this frame, we represent the Reading Process defining the following concepts:

- Reading: a sequence of Session(s) of the same work (first or one of multiple reading), sharing a context, such as being part of the same activities, addressing the same aims or approached with an overall state of mind
- Session: a sequence of Experience(s), an interaction between event and action delimited by the opening and closing of a reading resource

- Experience a change of state in a Session

Reading, Session and Experience are specialisation of Reading Process and specialisation of Temporal Entity. Therefore, one Reading can precede, follow or co-occur with a second Reading, and Reading Sessions and Reading Experiences can be sequentially ordered.

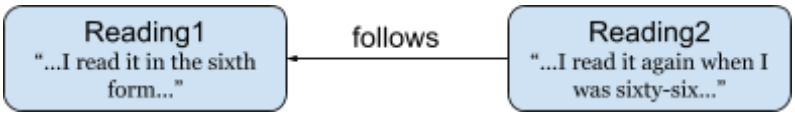
In this frame, Experience is necessarily and strictly *part of* one specific Session which is necessarily and strictly *part of* one specific Reading. On the other hand, a specific Reading instance provides the *situation* in which the multiple instances of Session occur, while each individual Session provide the *situation* in which the multiple instances of Experience occur.



The schema encoding Experience, Session and Reading as specialisation of Reading Process, and therefore of Temporal Entity.

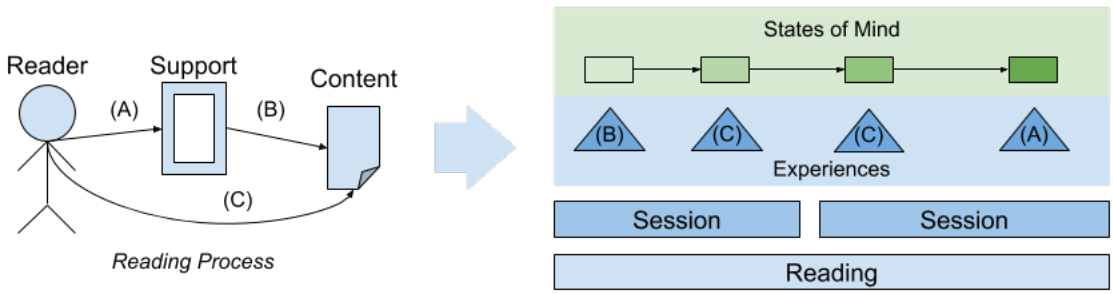
Example including information about temporal relations between readings:

> *"I was reading the Dostoyevsky again, I read it in the sixth form, and this is another aside, but I was very taken with Raskolnikov, you see, when I was eighteen. I thought he was a bit of a, you know. I read it again when I was sixty-six..."* Extract from Memories of fiction, Brenda C. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/brenda-part-2.

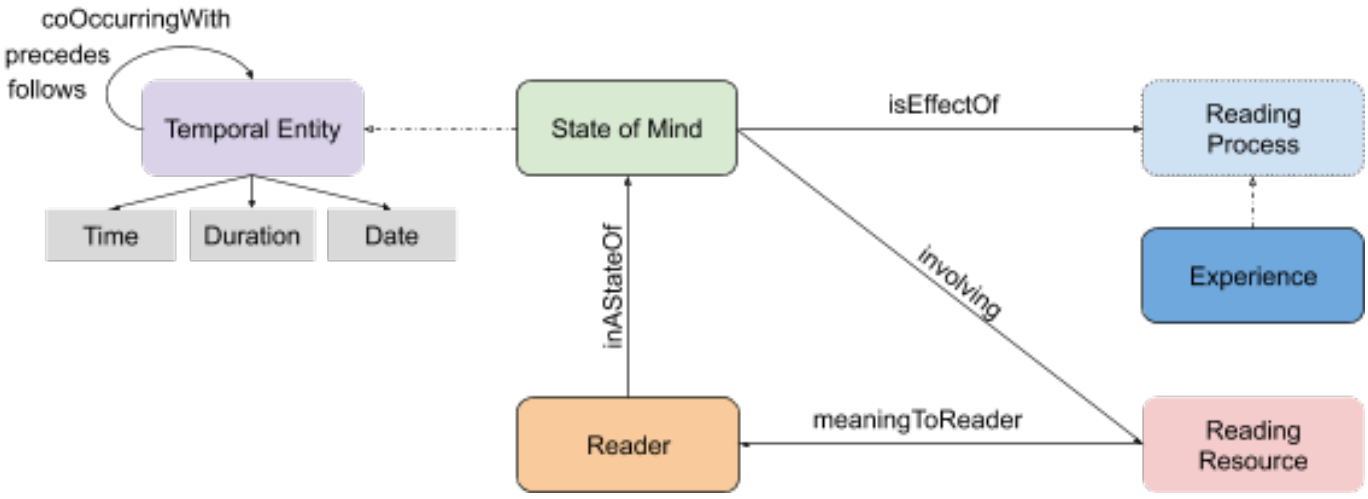


Experience of Reading

We consider the Experience an evidence of a change of the reader's State of Mind, triggered by the interactions a Reading Resource, a Content or a Medium. In general, an Experience could be related to the interaction with the Medium, Content or the combination of the two.



The interactions of reading (left) performed during reading Session may result in reading Experiences, and therefore in a change of the Reader's State of Mind.

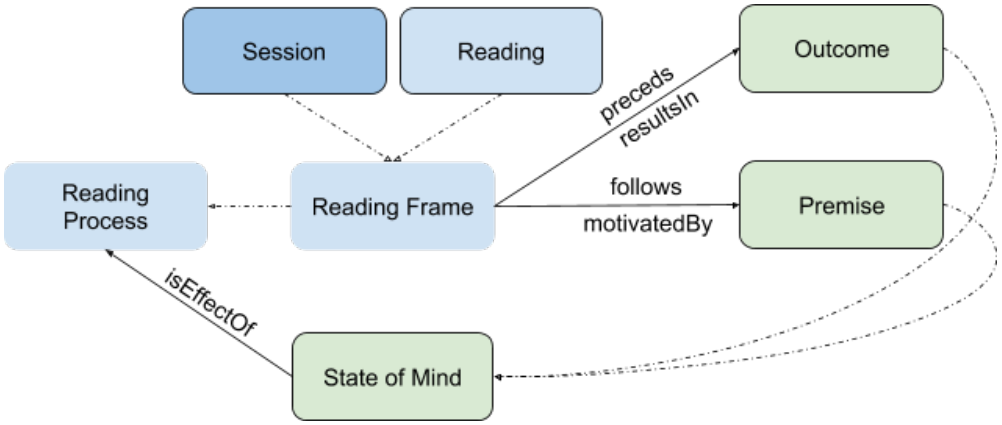


Schema representing State of Mind, specialisation of Temporal Entity, and its relations with Reader, Reading Process and Reading Resource

The State of Mind is a Temporal Entity. Thus, we can represent sequences of State of Mind, multiple co-occurring States of Mind or define temporal relations between State of Mind and other Temporal Entities, e.g. a State of Mind *following* an Event or *preceding* an Alteration of a Medium (before taking a note or ripping off a page from a book).

The states of mind occurring outside the scope of Reading or Session may have a special role in understanding a reading experience. For instance, a State of Mind can be related to the aim of reading or to a major achievement related to the reading of a chapter of a book. To address this case, we specify two subtypes of State of Mind: Premise and Outcome. We call Premise the State of Mind which precede a Reading or Session providing a *motivation* of the engagement of the reader in reading. We call Outcome the State of Mind *following* and *result* of a Session or Reading.

A State of Mind, co-occurring during a reading, is *effect of* a reading Experience, and therefore of the Session and the Reading. Differently from the State of Mind co-occurring during Sessions, the State of Mind occurring before or after Sessions or Reading are not strictly connected to a specific Experience. We take into account these cases by connecting these outer State of Mind to the Reading Frame (generalisation of Session and Reading). The relations between Outcome and Premise are characterised considering Session and Reading (Reading Frame) as an activity that the reader firstly approaches, in a specific State of Mind (Premise), and then lefts (or ends), with a new State of Mind (Outcome).

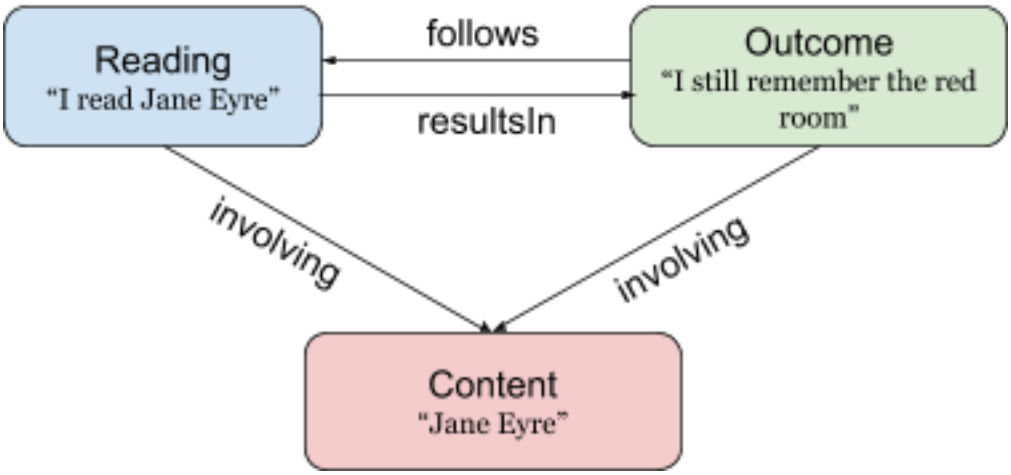


Schema representing Reading Frame, specialisation of Reading Process, and its relations with Outcome and Premise, specialisation of State of Mind, and Session and Reading

Considering sequences of Session and multiple instances of Reading of the same content, the Outcome of a Session of Reading can act as the Premise of the following Session or Reading.

Example including information about a state of minds following a reading:

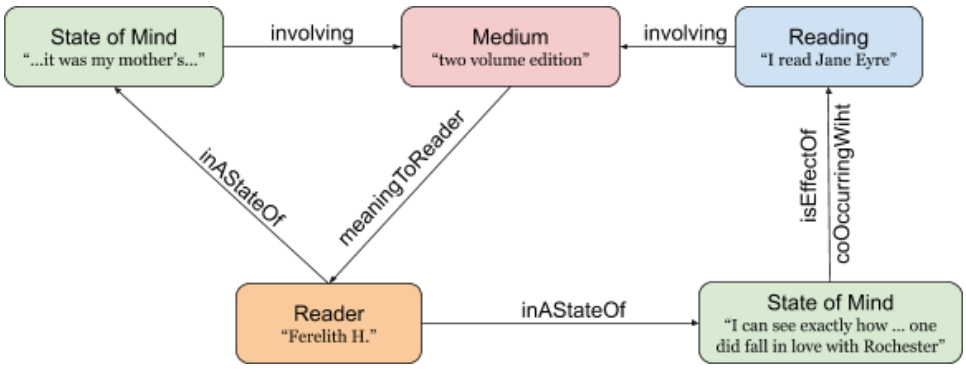
> *"... I read* Jane Eyre *... I still remember the red room ..."* Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.



From the analysis of sources, we identified multiple types or features of a State of Mind. Indeed, we found examples of descriptions of State of Mind about an *emotion* triggered by reading, a *judgment* about a passage, chapter or a content as whole, memories (*remembrance*) of specific words or events in the story, *self-reflection* about to the identification of the reader with a character. We take into account these evidence by providing a preliminary open list of subtypes of State of Mind.

Example including information about different types of state of minds related to reading:

> *"... I read ... my mother's two volume edition ... seemed to me rather an old fashioned book ... I can see exactly how ... one did fall in love with Rochester ..."* Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.

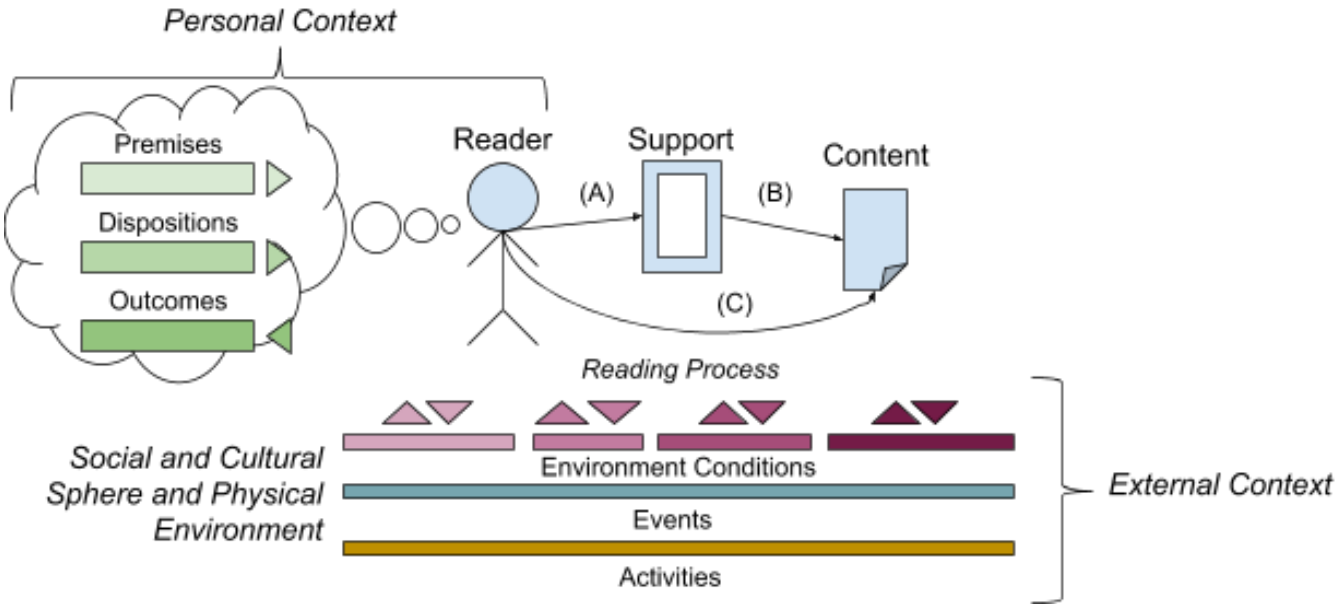


Lastly, a State of Mind can be either oriented toward the self or the social sphere. E.g. the *Aim* of passing an exam is *oriented* toward the reader's role as student, while the *Aim* of self-improvement is *oriented* toward the

achieving an "invisible" effect concerning mostly the reader condition (its essence). The orientation of a State of Mind is taken into account by introducing the *Orientation* property, which could be either: *Undefined, Autocentric* (concerning the self) or *Allocentric* (concerning the others).

The Context of Reading

Reading is embodied into a material and social context, and external context, and into a personal context including the reader's aims, expectations, plans, physical conditions, emotions, etc. We represent the external context with the use of the concept of Event, while the personal context, partially addressed through the concepts of Reader, Premise and Outcome by introducing a new subtype of State of Mind: the readers' Disposition.

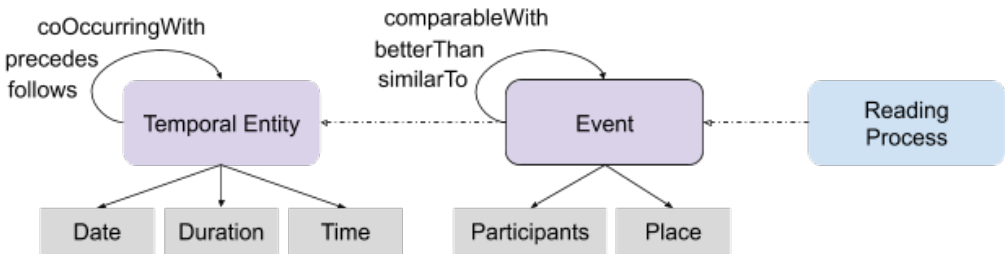


A Reading Process is embodied in a external and a personal context.

The External Context of Reading

An Event represents a state in a cultural, social sphere or physical environment (http://www.cidoc-crm.org/entity/e5-event/version-6.2). An Event is characterised by spatial-temporal features, and features related to the cultural and social systems involved, such as participants and place. As discussed in the introduction, we consider reading as an event as well. Thus, a Reading Process is defined as a specialisation of Event, sharing the same properties.

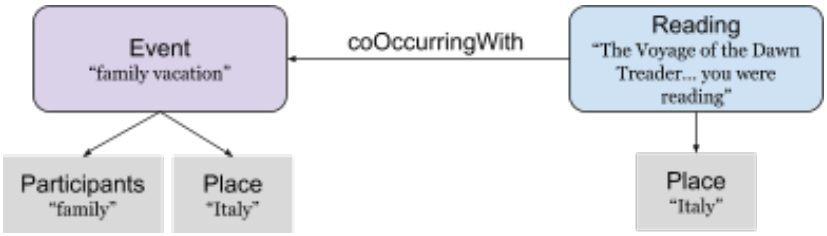
In the context of representing experience, the description of experiences by comparison is a practice emerging from the study of sources, e.g. the first Harry Potter was better than the second one. In the mode, we consider the comparison between different reading experiences as a properties of Event in general, and thus of reading.



Schema representing Event, specialisation of Temporal Entity, and superclass of Reading Process

Example of reading experience including information about a co-occurring event, family vacation.

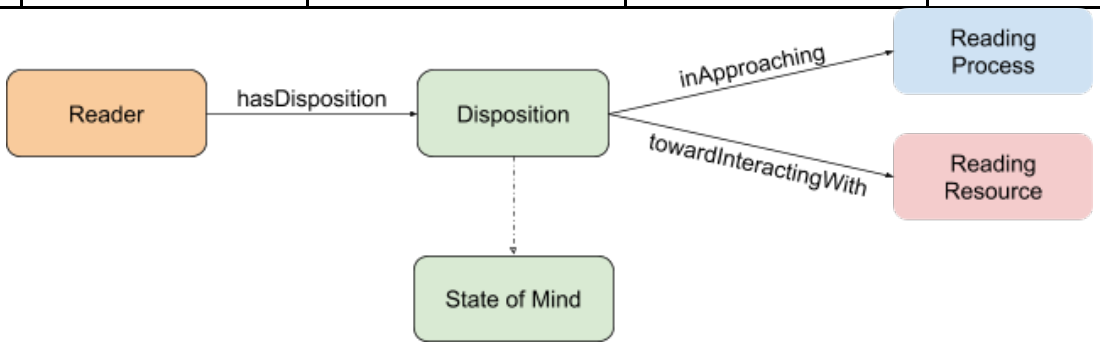
> *"... I think one found that throughout the whole of that sequence, The Voyage of the Dawn Treader... you were reading that in Italy, was it? Yes. On holiday with family. With family, yes."* Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.



The Personal Context of Reading

A Disposition is a State of Mind about *approaching* a specific Reading Process, in relation to a Medium, Content or both. A Disposition could be related to specific features of the Reader, such as linguistic skills, occupation, education level. We take into account the existence of different types of dispositions by providing a preliminary open list of subtypes.

Involving	**Examples of Readers' Dispositions**			
	Skills Disposition *e.g. Language, Age, Education, ...*	***Group Disposition*** *e.g. Nationality, Political orientation, ...*	***Aesthetic Disposition*** *e.g. Genre tastes, ...*	***Ethical Disposition*** *e.g. environmental concerns*
Content	the reader likes french language	Author of the same nationality	Favorite genre	Controversial topic
Medium	the reader never used before an e-reader	Old fashioned reader that likes printed books	Small typeface, difficult to read	Book in recycled paper
Content & Medium			New edition with revised translation	



Schema representing Disposition, specialisation of State of Mind, and its relations with Reading Process, Reading Resource and the Reader

READ-IT Data Model - Reading Experience

In this section, we provide an outline of the module of READ-IT data model about the reading experience. Following, we provide a preliminary discussion of the relations between this data model and relevant projects, with particular attention to FRBRoo and CIDOC-CRM. The formal encoding of the module will be object of a specific technical report.

As general premise to this module, we expect to import from FRBRoo, CIDOC-CRM and FoaF most of the classes of this module, specifically: Temporal Entity, Reading Process, Reading Frame, Event, Person, Content, Medium and Alteration (see the section about alignments). The concepts that we do expect to implement are the concepts strictly related to the reading experience: Reader, State of Mind, Reading and Session. In the first case, we refer to the documentation of the these projects, while in the second case we refer to concepts which are still object of research. Thus, for different reasons, this document provides just an outline.

A secondary premise is related to two general requirements about the READ-IT data.

Firstly, we have a technical requirement related to the training machine learning algorithms using READ-IT data. Specifically, this technical requirement is about making data explicit as much as possible. This requirement is taken into account by providing relations which could be derived so that it will be possible to explicit such derivations. For instance, a Reader is *engaged in* a Reading, and *reading* a Content, we provide a relation *involving content* so that we can explicit that the Reading is *involving content* Content. Even though this relation could be easily inferred from a semantic graph, the training of machine learning algorithms require explicit facts.

Last but not least, there is a requirement related to the use in research of data generated through the use of the model of reading. Specifically, the result of the analysis of sources of reading experience should be encoded preserving the formulation provided. Let's consider the following extract:

> *"... The Voyage of the Dawn Treader... you were reading that in Italy, was it? Yes. On holiday with family ..."* Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.

This extract can be encoded at least as:

(a) Reader (implicit) is *reading* a Content (The Voyage of the Dawn Treader) in a *Place* (Italy) with *Participants* (Family)

(b) There is a Reading, *Location* Italy, *co-occurring with* an Event (holiday with family), *Location* Italy, and *involving* a Content (The Voyage of the Dawn Treader)

(c) Reader (implicit) is *reading through* a Medium (implicit) *providing access to* Content (The Voyage of the Dawn Treader). The Reader *is engaged in* a Reading has *Location* (Italy) and it is *co-occurring with* an Event (holiday with family in Italy).

These three different encodings highlight that: firstly, there is an element of subjectivity in the interpretation of the reading experience, and secondly the encoding can be introduce more to what the source says. Each possible encoding point out different issues:

- In (a), the Reader is not mentioned in the source. Thus, encoding a Reader is not wrong but it does not represent the source. The existence of a Reader should be derived by the existence of a Reading. Similarly, in (c), we encode a Medium that is not mentioned in the source.

- In (a), the family is considered as participants in the reading. This could be related to prior knowledge of the annotator, for instance about the reader or the context of the sources (experiences of reading to others).

- Some concepts could be considered marginal by the annotator, e.g. the holiday as Event, see differences between (a) and (b)

- In (b). the annotator can include his deduction about the location of the holiday, Event, through deduction, the reading is in Italy, and it co-occurs during the family holiday, Event, which therefore is in Italy as well. The Location of the event may be the result of a correct deduction, but it is not distinguishable from the Location of the reading which is explicitly stated in the source.

Summarising, we the requirement for the use of the generated data in research is to avoid hidden interpretations of data at the time of encoding, keeping the encoding adherent with the sources. The encoding of the extract should include all and only concepts explicitly reported in the source. But the encoding should still be able to represent as a reading experience, relating the set of included contents together. Practically speaking, we cannot rely on rigid pattern to relate concepts together.

Following the example, the encoding of the fragment should include the following facts:

- Reading:1 has *location,* value "Italy"

- Reading:1 *involve content* Content

- Content:1 has *title*, value "The Voyage of the Dawn Treader"

- Reading:1 *co-occurs with* Event

- Event:1 has *description*, value "holiday"

- Event:1 has *participants*, value "family"

It is worth noticing that regardless the introduction of the concepts of Reader, Medium or Session, this fragment can be encoded, and all facts can be related each other.

The extraction of the implicit knowledge from the encodings of sources will be performed through logical inference. The facts derived through reasoning will be identifiable as such, preserving the primary encoding done by annotators. Considering the previous example, the final result of annotation and reasoning will be a sum of all three encodings, thus including facts grounded on sources and facts derived through reasoning.

The presented approach will configure the READ-IT database as a collection of multi-layer graphs and sources, representing the plurality of contributions and the several steps performed to enrich the representations of reading experiences. The result will be the mitigation of the different "styles" of encoding, without losing the trace of the direct connection with the source created by annotators. In this frame, researchers could decide whether to use their own data, strictly as generated by them or to include the result of inferences or data generated by other human or automatic annotation of sources.

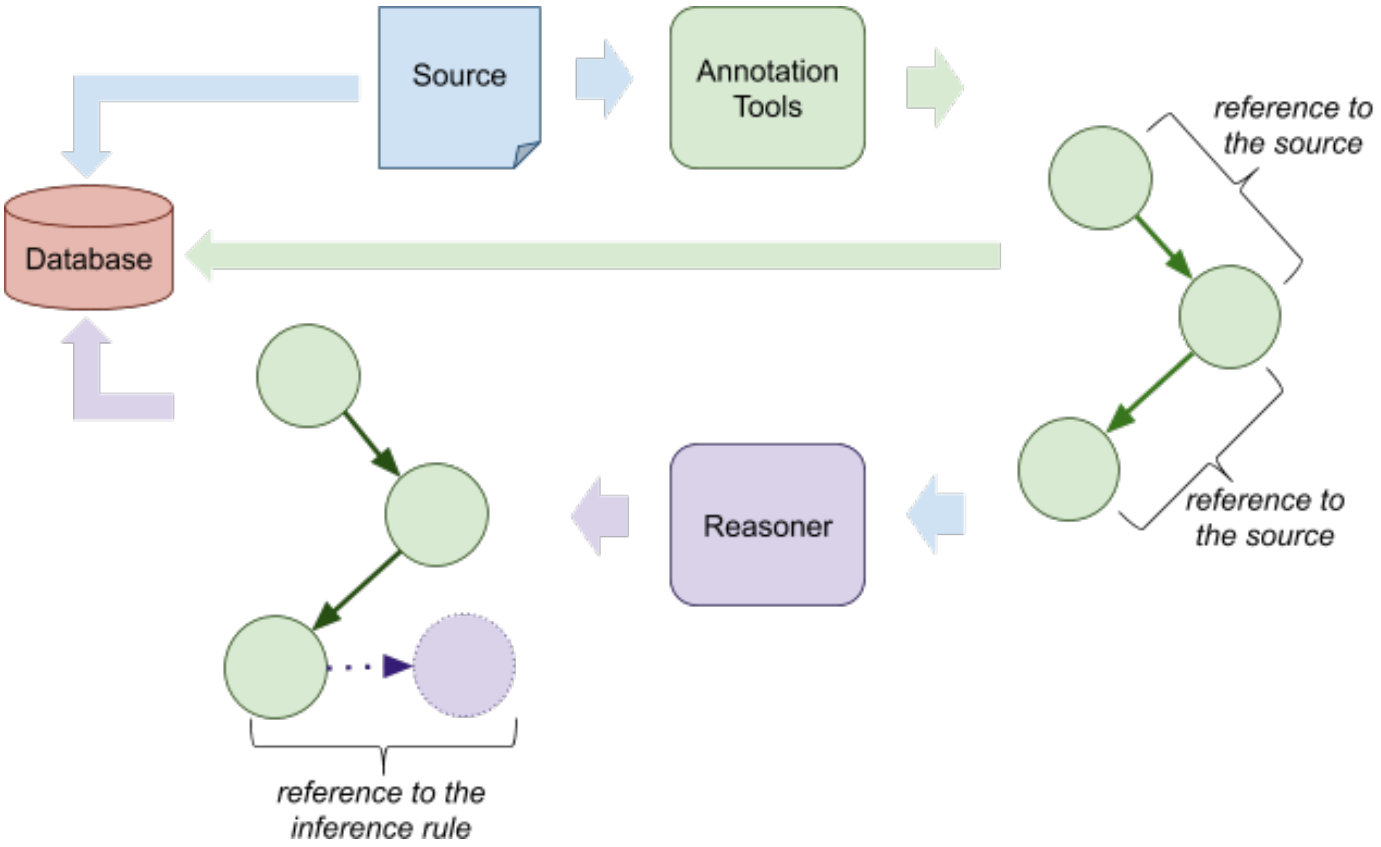


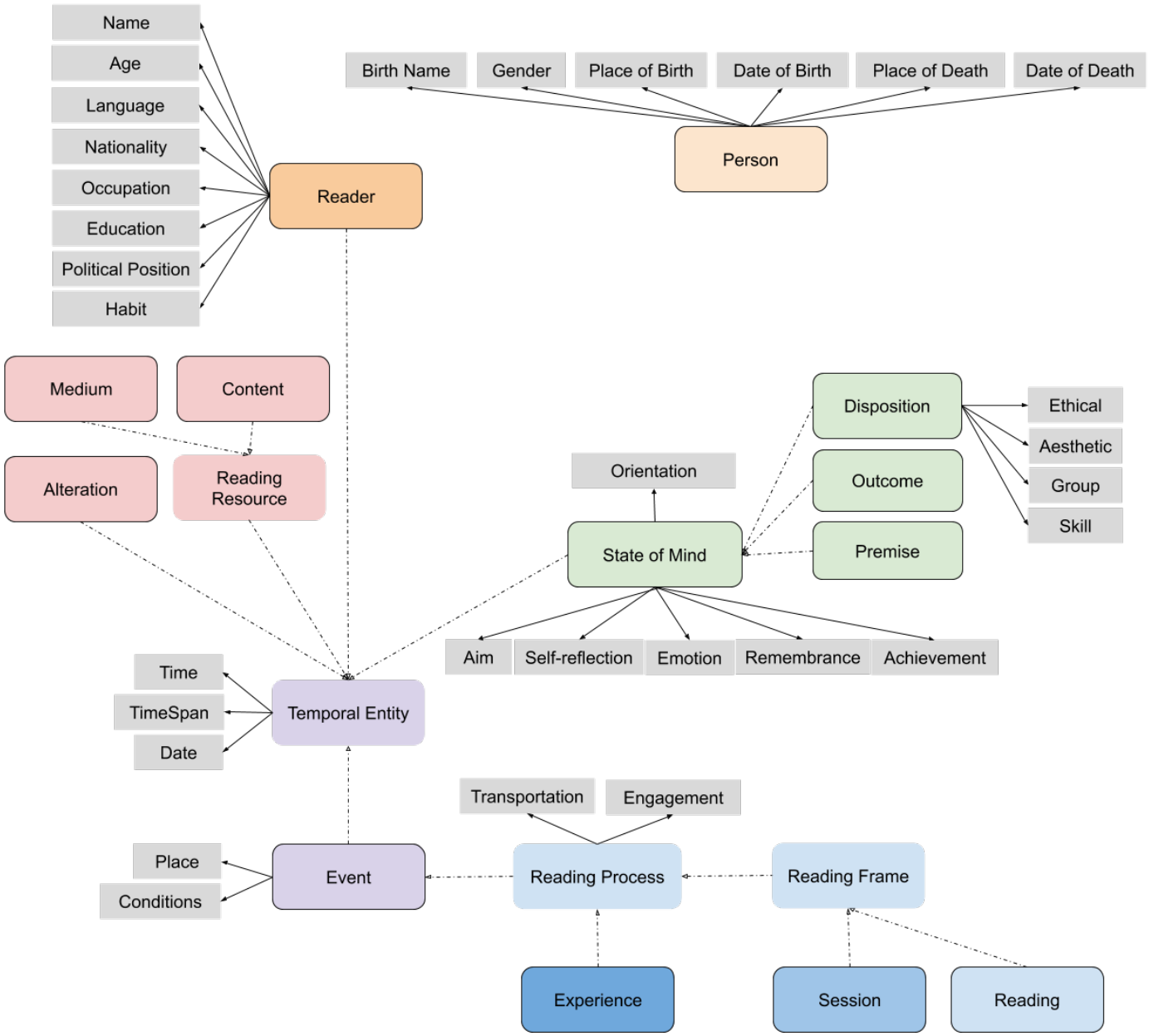
Diagram representing the distinctions between annotation and inference steps in the construction of the representation of reading of reading experiences from sources.

The Model of Reading - V1.6

Classes of READ-IT data model

Legend:

- Boxed round corners represent classes
    - Continuous border boxes, represent instantiable classes
    - Boxes without border, represent non-instantiable abstract classes
- Arrows represent relations
    - dotted arrows indicates *specialisation* relations between classes
    - unlabelled continuous arrows indicate class properties
- Sharp-edge borderless boxes indicates class properties
- Colours indicate a thematic area of the class, to facilitate the reading of the diagram



It is worth mention that the model as described includes abstract classes: Reading Resource, Temporal Entity, Reading Process and Reading Frame. These classes are not meant to be instantiated, i.e. we do not expect to use them to represent reading experience. We use these classes as generalisations and abstractions of other concepts exploiting the inheritance of properties and relations between classes (from general to specialised classes).

Lastly, to keep the focus on the reading experience and avoiding making assumptions on open issues, the definition of several properties of a class, e.g. nationality, habit, education, is a simplification of more complex design patterns. In the future development of the data model, these properties will be replaced by specific classes and relations.

Relations of READ-IT data model

Legend:

- Boxed round corners represent classes
  - Continuous border boxes, represent instantiable classes
  - Boxes without border, represent non-instantiable abstract classes
- Labelled continuous arrows indicate (oriented and symmetric) relations between classes
- Colours indicate a thematic area of the class, to facilitate the reading of the diagram



As mentioned at the beginning of this section, as design principle we wish to support multiple patterns for representing reading experiences. An important aspect is to define a comprehensive set of relations including inverse functions, e.g. *better than* inverse of *worse than*, *precede* inverse of *follows*, symmetric relations, e.g.

comparable with, co-occurring with and alternative paths, generalisations as specialisations, such as Reading Process *involving* Reading Resource, *involving medium* Medium, Reader *in a state of* State of Mind *involving* Medium which has *meaning to reader* Reader.

The presented list of relations, in the diagram is not complete, but it will be integrated in the formal encoding of this module of the data model.

Alignment with Standards and Resources

In its formal implementation, the READ-IT model of reading will make use of existing standards, design patterns and conceptualisation, such as: *Foaf, FRBR, CIDOC-CRM, FRBRoo and Schema.org.* Following, we provide a **preliminary discussion of the alignments** between READ-IT data model and FRBRoo, CIDOC CRM, FoaF and Schema.org. It is worth to point out that **we do expect major revisions** to the alignments here discussed.

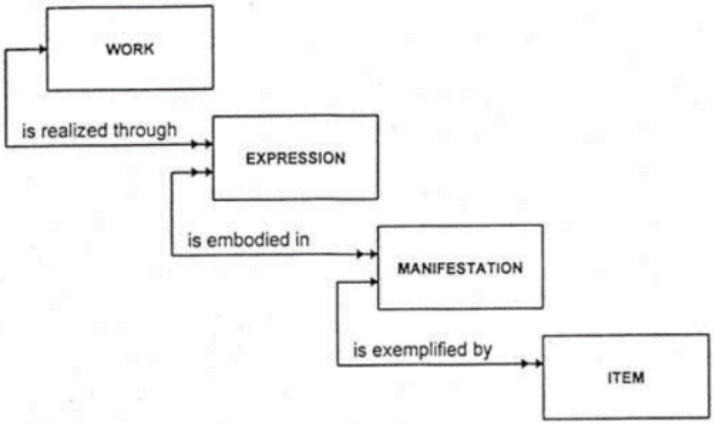
Friend of a Friend (FoaF) is an dictionary to describe web resources *"to help us tell the Web about the connections between the things that matter to us"* (http://www.foaf-project.org/original-intro). Specifically, we consider the class *Person* of FoaF as general description of the reader's features.

> *"The Person class represents people. Something is a Person if it is a person. We don't nitpic about whether they're alive, dead, real, or imaginary. The Person class is a sub-class of the Agent class, since all people are considered 'agents' in FOAF."* (http://xmlns.com/foaf/spec/#term_Person).

Schema.org (Schema) is a project with the ambition of providing schema (vocabulary and entities) to describe web resources. Schema.org follows a top-down approach but it is based on the experience of the web community and organisations such as Google, Microsoft, Yahoo and Yandex (founders of Schema.org). Schema.org provides description of a wide range of web resources of relevance for READ-IT, such as:

- *CreativeWork* (https://schema.org/CreativeWork), provides an description of web contents, e.g. blog posts, magazines, web novels and in general almost any possible content. It should be used as specialisation of *Content* as fallback in case the FRBR entities could not be applied
- *ReadAction* (https://schema.org/ReadAction) *"the act of consuming written content"*, it characterises the relations between the reader, Support and Content as superclass of *Reading Event*
- *UserInteraction* (https://schema.org/UserInteraction) *"… way of talking about users interacting with pages"*, it describes the contributions on a page and the public evaluation providing the context for the interaction. It is used as specialisation of *Content.*
- *Event* (https://schema.org/Event) *"An event happening at a certain time and location, such as a concert, lecture, or festival." Event* is the superclass of our READ-IT *Event.* In case of group reading, the *performer* is the reader reading aloud, while the *participants* are the readers listening

Functional Requirements for Bibliographic Records (FRBR) is the current standard for the representation of bibliographic resources. FRBR provides a conceptual framework based on four main concepts: *work*, *expression*, *manifestation* and *item* (IFLA 2009). This conceptualisation is of great importance for READ-IT for two reasons: 1) it is a standard for bibliographic record and as such, it provides the possibility to connect READ-IT annotations to these catalogues, and 2) this conceptualisation provides a finer grained distinction between Content and Support.

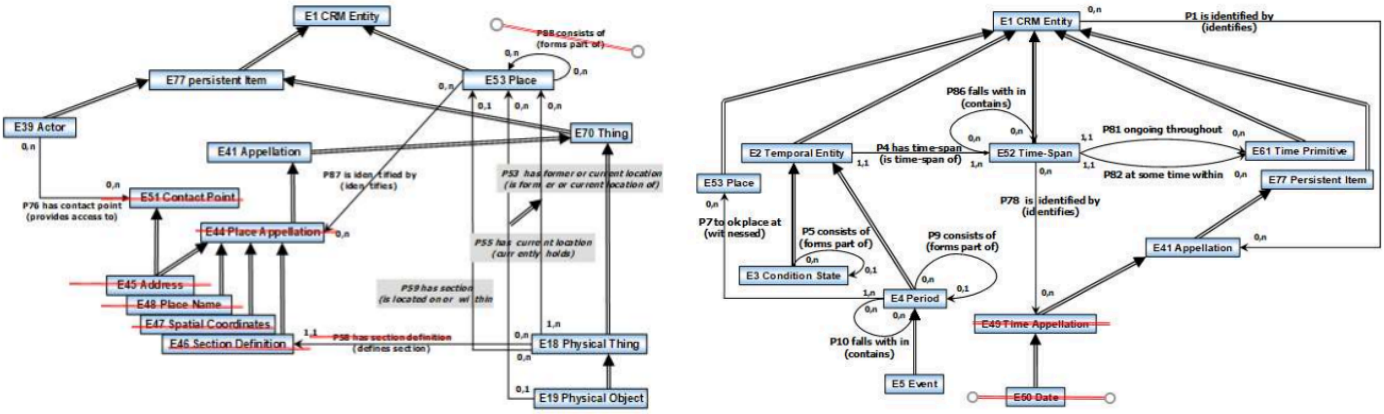


Source https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf page 14.

The CIDOC Conceptual Reference Model (CIDOC CRM) is an official standard, ISO 21127:2006, providing general concepts and an ecosystem of specialist ontologies for representing cultural heritage resources. From the CIDOC-CRM website (http://www.cidoc-crm.org)

> "The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation."
> *Source "What is the CIDOC-CRM?" http://www.cidoc-crm.org.*

The CIDOC-CRM focus are cultural heritage documentations, such as a relevant set of sources object of study in READ-IT case studies. Thus, the alignment or adoption of CIDOC-CRM concepts is a natural choice.



CIDOC-CRM general concepts and recent changes in version 6.2.6. Souce http://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v6.2.6_Definition_esIP.pdf

Reader & Person

In our data model, the Reader and Person is a specialisation of FoaF Person, and in general an instance of a FoaF Person could be connected with multiple instances of Reader by a *description of* relation. An instance of FoaF Person is the static description of the person reader, while the instances of Reader are the description of the person at the time of the reading.

In READ-IT data model, we consider FoaF:Person and Schema:Person superclass of Reader and Person, their features can be used to describe both of them. But, as argued, the variant features should be used to describe the reader, while the invariant features to describe the Person. In is worth to highlight that FoaF:Person and Schema:Person introduce reflexive relations, such as FoaF:*knows* that can be used to represent the network of relationship among readers and persons.

FoaF:Person	Schema:Person	Reader	Person
name	givenName		Name
surname, lastName, family_name	familyName		
gender	gender	X	X
birthday	birthDate		X
knows	colleague, knows, parent	X	
workplaceHomepage, workInfoHomepage, currentProject	hasOccupation, jobTitle, worksFor	Occupation	
publications	memberOf	X	Education, Political Position, Occupation
schoolHomepage	knowsAbout, alumnyOf	X	Education
…			

The relations between the properties of FoaF:Person, Reader and Person could be of specialisation or potential relevance from FoaF toward READ-IT or the other way around. For instance, *publications* could inform about either *Education, Occupation* or *Political Position* of the Reader.

Medium & Content

In our data model we provide the concept of Content and Support. The FRBR equivalent of the Content are *work* and *expression*, representing the creative work and its realisation through one specific form (expression). The *Support* encompass the concepts of *manifestation* and *item*, the first is the form of the physical embodiment of the *expression* (how the expression looks like in a physical form), e.g. the first edition of *La Montagna Incantata*, translation of Ervino Pocar published by Acquerelli, while the *item* is one of the instances of the *manifestation*, e.g. my personal copy stolen to my father.

Alignment with FRBR

- *Support* = *item* if the reader refers to the specific object
- *Support* = *manifestation* if the reader refers to the technical features of the support
- *Content* = *expression* if the reader refers to the literary form of the content, the translation, illustrations
- *Content* = *work* if the reader refers to the author's ideas, the message, the story, etc.

In this sense, we consider *Support* superclass of *item* and *manifestation*, and *Content* is a superclass of *expression* and *work*. This first set of entities (Group 1 Entities and Primary Relationships, *work*, *expression*, *manifestation* and *item*) is one of three groups of subjects of a *work*. These three groups of entities are of relevance in representing reading experiences, since providing an explicit connection (*has a subject*) between the Content (FRBR:work) and a vast range of entities, see next figure.

In this regard, we could annotate the relations between the experience of Readers and *concepts*, *objects*, *event* or *places* in terms of causes of a specific aspect to of the experience, e.g. a *concept* addressed in a content or the *corporate body* publishing the content could be the reason for the *Reader*'s *Dispositions* toward the content.

Thus, as a Content can be related to other entities, we can implicate these entities within the Experience(s) and Outcome(s) of Reading, and with the relative changes of State of Mind and Disposition(s) of the Reader. The relation of FRBR *has as subject* has an equivalent in the propertoy of Schema:*CreativeWork* *about*.

Aside from to providing a relation to describe the topic of a Content, Schema:*CreativeWork*, provides a general description of web contents, e.g. blog posts, magazines, web novels and in general almost any possible content. It should be used as specialisation of *Content* as fallback in case the FRBR entities could not be applied.

FRBR	Schema:CreativeWork	Content	Support	Alteration
title of the work	name	X		
has as subject	about	X		
	accessMode		X	
	genre	X		
	inLanguage	X		
	correction			X
	comment			X
Publication Status		X		
Title of the Expression		X		
…				

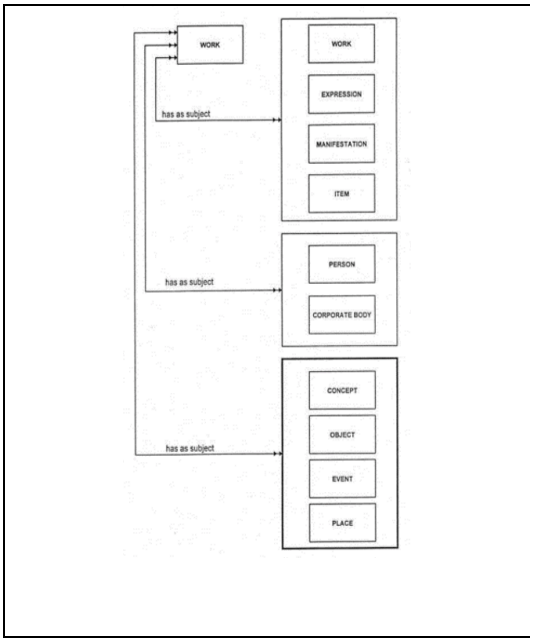
Furthermore, Schema:Thing and FRBR *item* provide an extensive set of properties for the Support and Alteration.

FRBR	Schema:Thing	Support	Alteration
title of the item	name	X	
access restrictions on the item	accessMode	X	
condition of the item			X
marks/inscriptions			X
…			

FRBR provides a set of relations to describe responsibilities in *work*, *expression*, *manifestation* and *item*. From FRBR definitions (IFLA2009) Entities and "Responsibility" Relationships, there are two possible actors bearing responsibility:

- *person*: an individual
- *corporate body*: an organization or group of individuals and/or organizations

The relations of responsibility of *person* and *corporate body* and describe the specific role of these actors respect a *work*, *expression*, *manifestation* or *item*, which could be of relevance for the Reader's disposition toward Content and Support. For instance, a Reader could have a consolidated trust on an editor about publications on a specific topic.



On the left, the three groups of concepts subject of a work. On the right, the group 2 Entities and "Responsibility" Relationships from FRBR documentation (IFLA2009, page 15). Source FRBR (IFLA2009) (https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf).

In READ-IT we adopt the class FRBR:corporate body and by considering FRBR:item a superclass of Support and FRBR:expression, FRBR:manifestation, FRBR:work superclasses of Content. Thus, it is possible to use the relations *is owned by, is produced by, is realized by* and *is created by* to describe an extra set of responsibility-related relations.

States of Reading

In READ-IT there are two concepts representing the state of an object:

1. the Reader representing a state of the Person
2. the State of Mind of the Reader, characterising the Reader cognitive state

Furthermore, the concept of Alteration of a Medium refers to an implicit notion of state of the medium.

In this view, CIDOC-CRM provides the concept of condition state, E3_CONDITION_STATE, addressing "the states of objects by a certain condition over a time-span" (http://www.cidoc-crm.org/Entity/e3-condition-state/version-6.2).

Lastly, the concept of Tempora Entity of READ-IT seems to be equivalent to CIDOC-CRM E2_TEMPORAL_ENTITY.

Situation of Reading and CIDOC-CRM & Schema.org

The Reading Process represent in READ-IT the implementation of the action of reading. In this frame CIDOC-CRM provides the concept of event, E5_EVENT, and activity, E7_ACTIVITY, (specialisation of event). The concept of Event in READ-IT appears to a preliminary analysis compatible with E5_EVENT, and the concept of Reading Process (generalisation of Reading, Session and Experience) with E7_ACTIVITY.

Schema.org provides a set of properties related to "the act of consuming written content" within the concept of ReadAction. By adopting Schema:ReadAction as a superclass of Reading Process we can use such properties to describe the **situation of reading**, Reading Frame, and the **status of the process**. Furthermore, ReadAction provides properties to create relations between Reading Process and Reading Frame with Content, Reader. Lastly, as we consider reading as also an Event, ReadAction provides properties about *participant(s)* and *location.*

Schema:ReadAction	Reading Process	Reading Frame	Event
agent	Reader		
object	Content		
Instrument	Support		
participant			Reader, Person
startTime, endTime			Time, Date

About Event, Schema.org provides the concept of *UserInteraction* to provide an external perspective of an user action. Indeed, using Schema:UserInteraction as superclass of Event provides as consequence an extra set of

properties to describe Reading, Session and Experience as Event(s), such as the *location* of reading, the *attendee(s)*, *language.* Furthermore, the extension of Event provides the ability to describe co-*occurring* Events, taking place at the same time of Reading (s), Reading Session(s) and Reading Experience(s), and a set of relations connecting EventReaders.

Schema:UserInteraction	Reading, Session, Experience	Event	Reader
actor, audience, workPerformed	engagedIn		X
duration		X	
inLanguage		X	
location	X	X	
subEvent, superEvent	partOf		
startDate,endDate		Date	

Using the Data Model

READ-IT data model is a framework of concepts that could be used to structure the annotations of reading experiences. In this regard, using READ-IT data model is the process of structuring a source of reading evidences using the data model concepts, linking these concepts to fragments of evidences.

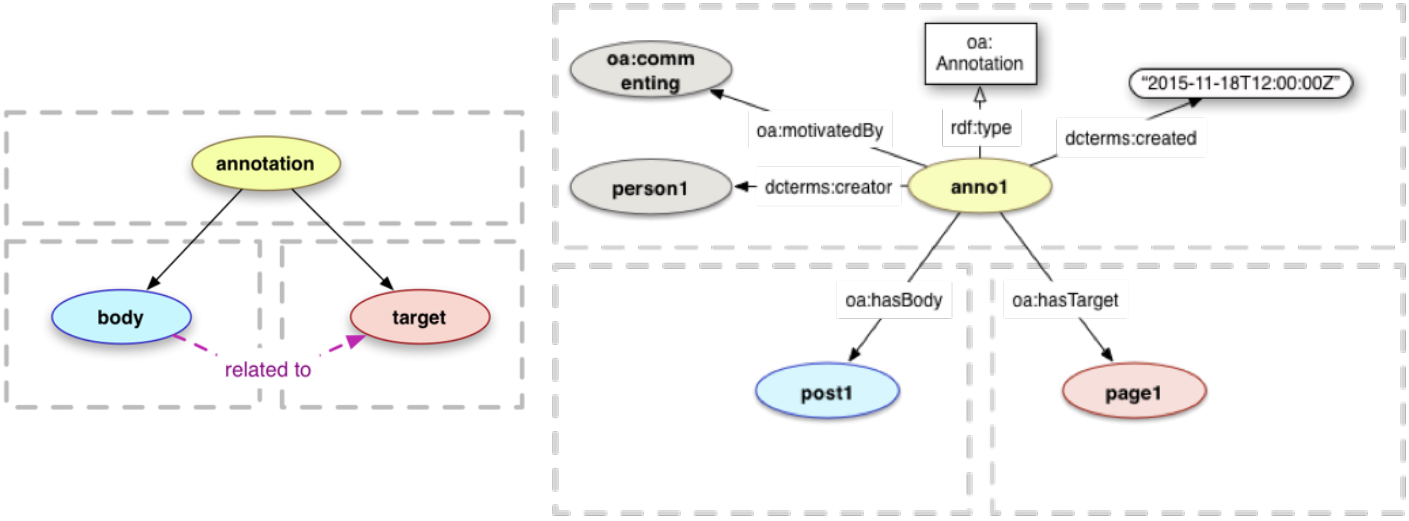
Annotations Metadata in READ-IT

An annotation is composed by a *body*, the conceptual structure build using the concepts of the data model, a *target* as fragment of the evidence, and a set of information about the generation of the annotation.

W3C Web Annotation Data Model

W3C Web Annotation Data Model: standard description model and format to enable annotations to be shared between systems.

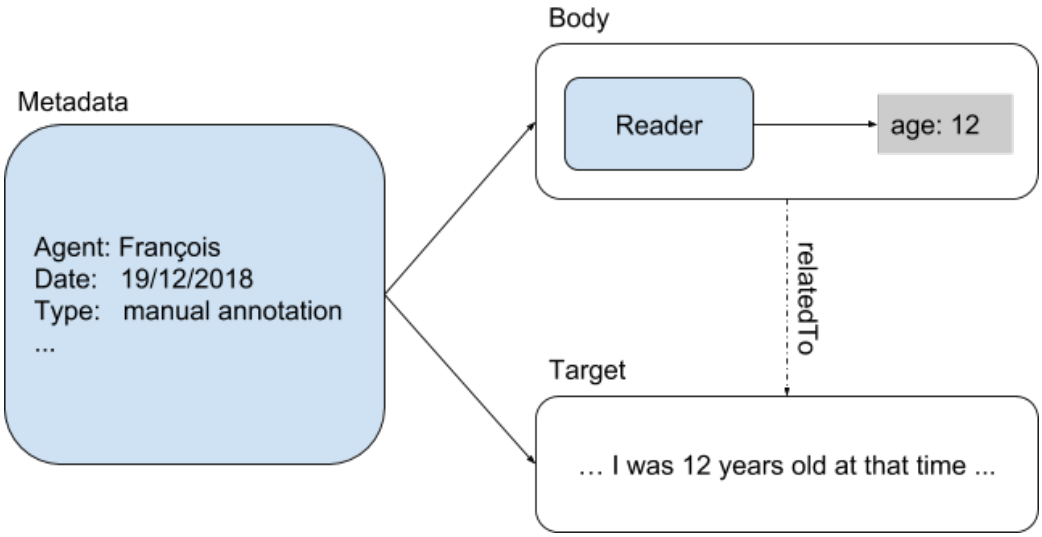
- *annotation* (evidence)
    A. *body* (data we collect analyzing)
    B. *target* (resource excerpt)
- The body is about the target, e.g. graph represents an interview



On the left, conceptual schema of W3C web annotation model. On the right, an example.
Source https://www.w3.org/TR/annotation-model/.

The annotation will point to the original source. Therefore, we will be able to extend the body after the data collection, e.g. we advance on the data modelling and analysis techniques and therefore we wish to extend the collected data.

Thus, the collection of annotations configures two different resources: a data and a metadata graph. The first includes the *conceptual structures* extracted from the reading experiences, while the second represents *how*, *where*, *who* generated the conceptual structure.



Consequently, with READ-IT database we refer to both the collection of annotation bodies (data graph) and the collection of metadata (metadata graph). The metadata graph includes all the information about the annotation with the references to sources object of the annotation (target) and body generated through the annotation processes with the reference to the source fragment.

Example 1

Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.

> *"Okay. Do you remember any of those? Do you remember Jane Eyre?*

Oh yes, oh, I remember Jane Eyre.

How old were you when you read Jane Eyre?

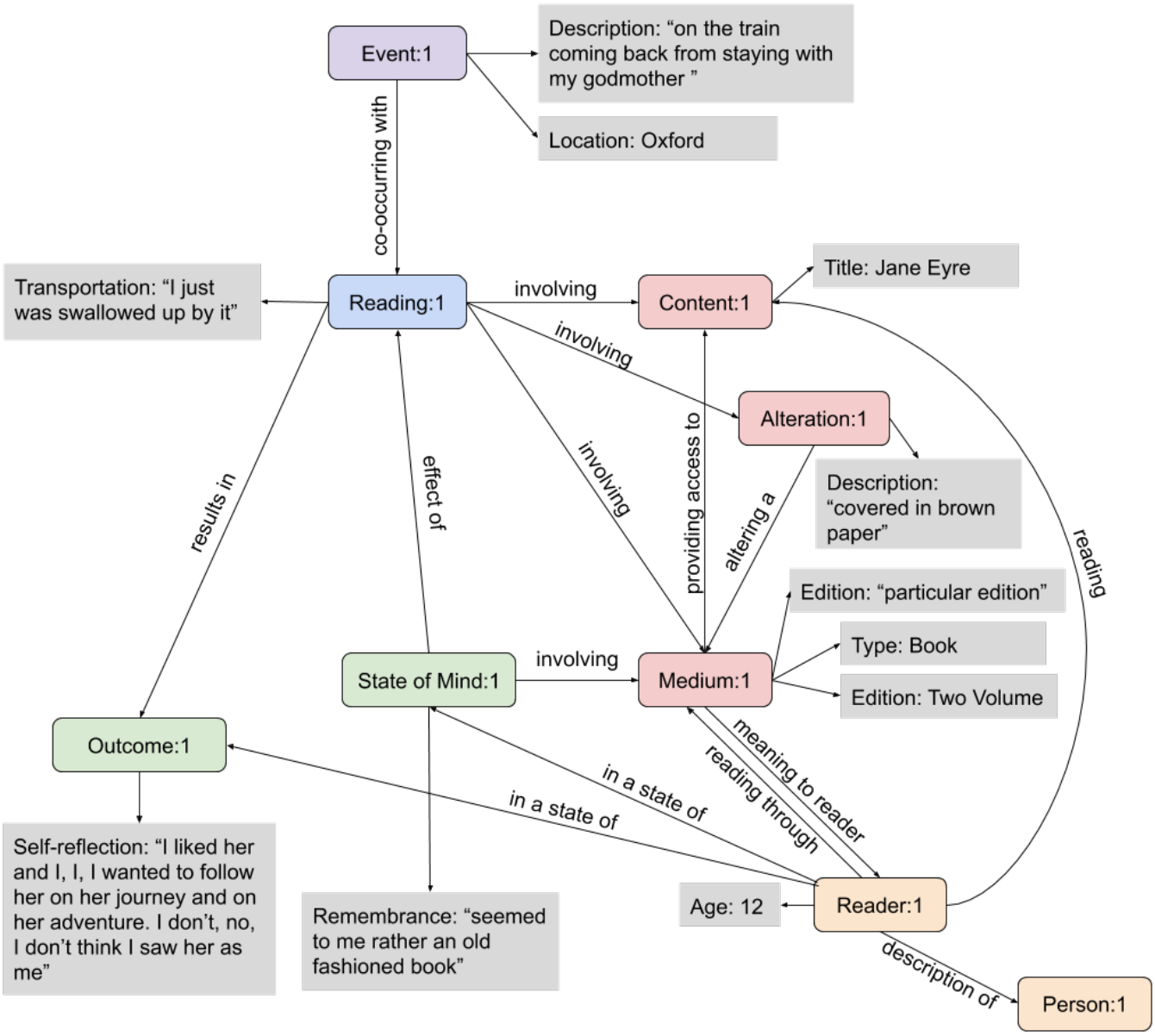
Well, I think I must have been about twelve. I, I have a very hazy memory of ages.

It's a long time ago.

But, ah, everything seems to either be at the age of eight or the age of twelve, but I'm pretty certain it would have been about twelve because I read Jane Eyre on the train coming back from staying with my godmother in, just outside Oxford, and I remember it was my mother's two volume edition and we'd covered it in brown paper, and I think she'd covered it in brown paper really to sort of make me feel better about reading rather, what seemed to me rather an old fashioned book. It didn't have a cover with pictures on it. It was bound, I think in, white vellum or so, it was, but it was a particular edition, and I remember, I just was swallowed up by it. I, I still remember the red room, just the, the fact, and the first meeting with Rochester, who, despite all the feminist arguments, and yes, I can well, I can see exactly how the, that the viewpoints, but one did fall in love with Rochester [laughter]. Why, why, I don't know, but, but one did.

But did you identify with Jane Eyre? You've been saying you didn't really identify particularly with the characters in, ah, The Lion, the Witch and the Wardrobe, was Jane Eyre, do you think, more of an identifiable character?

Yes, yes, to a certain extent, I think, but I liked her and I, I, I wanted to follow her on her journey and on her adventure. I don't, no, I don't think I saw her as me. I mean, I, she, but I, perhaps something, aspects that I would like to be that I, I did find her, I don't think I've ever really been the characters."



Simplified graph representation of Example 1. The features of Entities (e.g. Transportation) are fragments extracted from the source (e.g. "I just was swallowed up by it") and structured values represented with grey boxes (e.g. Title: Jane Eyre).

Example 2

Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.

"I think one found that throughout the whole of that sequence, The Voyage of the Dawn Treader, to discover that John Mandeville had written that this was the, the, later, to discover that these were stories that people in the Middle Ages, he was playing around with stories that already existed. I, I, for me that's, that's always been an interest and an, an important part of what makes a book really good.

And you said you were reading that in Italy, was it?

Yes.

On holiday with family.

With family, yes.

Em, and you read it a chapter at a time. And do you remember anything about the actual book, the cover, the smell, anything?

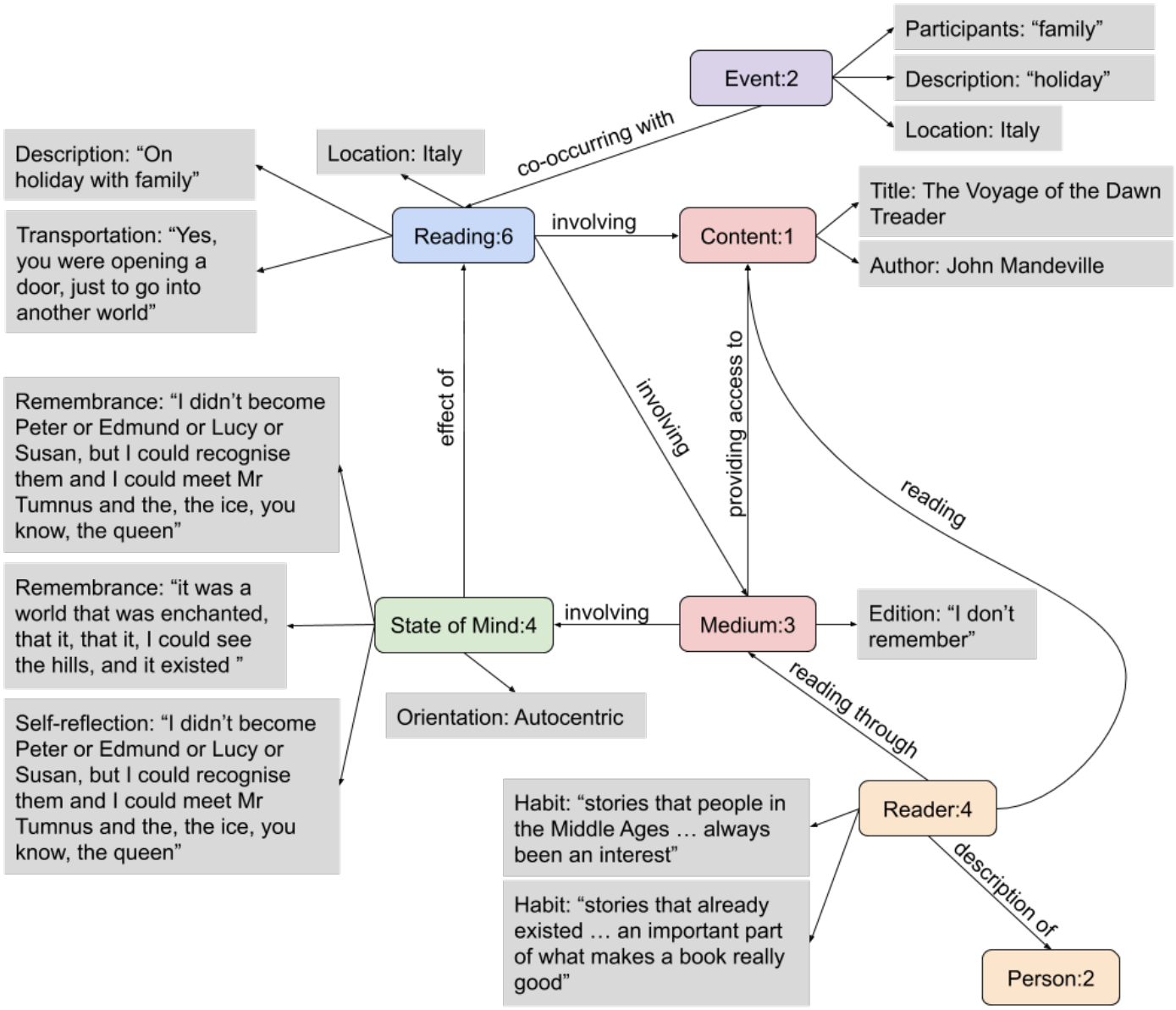
No, I don't remember. It would have been, it would have been the absolute, ah, you, it would have been the hardback edition with [Pauline Baines 00.10.27] because, as far as I know, this would have been '57, '58, I don't know, I can't remember when it was published, but it must have been very soon after, so I mean, I'd have to look up and find out, but as far as I know, this, it would be very close to its original publication date and paperbacks didn't really exist, so, not as such, so, well, a little bit, but it would be, it would have been the hardback edition.

And do you remember how it made you feel, reading it? Excited or … ?

Yes, you were opening a door, just to go into another world, which I always have loved. I, I mean that's, I wasn't terribly keen on the present day [laughter] generally, but it's this, just completely going into another world and living …

Yeah, opening that door and …

Living another life in a sense, or following in the footsteps of, I don't remember particularly identifying with any of the children as such. I didn't become Peter or Edmund or Lucy or Susan, but I could recognise them and I could meet Mr Tumnus and the, the ice, you know, the queen, ah, and it was a world that was enchanted, that it, that it, I could see the hills, and it existed …"



Simplified graph representation of Example 2. The features of Entities (e.g. Transportation) are fragments extracted from the source (e.g. "I just was swallowed up by it") and structured values represented with grey boxes (e.g. Location: Italy).

Design Patterns

As part of the design methodology we engaged in encoding a set of "significant" examples of reading experiences. This type of exercise was foundational in the design of the data model and to identify a set of recurrent patterns. In this section we present these recurrent patterns and how the data model should be used to encode them. These design pattern provides an *in-out* outline of the READ-IT data model, a measure of the complexity of the situations that could be encoded with it.

Reading to an audience

A group of people is reading together a chapter of book, one reading and the rest of them listening. This scenario is quite common, it could be reading a bed story, a reading group or a presentation event. In this scenario, the Reader is one actually *reading through* the Medium, while the listeners listening the reading aloud are not, strictly speaking, readers.

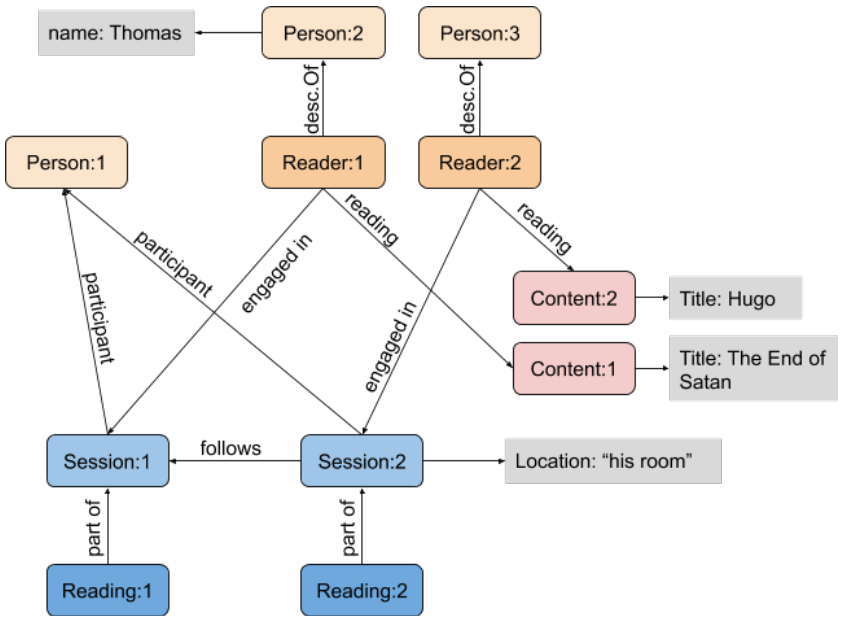
In this frame, we have a Session *co-occurring* with an Event. The Session is also an *Event* with *participants* listening to a performer, the Reader *engaged in* the Session.

Extract from Jean Lambert, Journal 1940-1942. http://publications.univ-provence.fr/ddb/document.php?id=98

> *"Il est revenu un peu plus tard, au moment où Thomas me lisait les passages de La Fin de Satan dont il m'avait parlé pendant la route, et en particulier le fameux vers:*
>
> *Baleine à la mâchoire infecte et délabrée... [Dieu, II, 7]*
>
> *qu'il déclamait en imitant Gide - et j'ai failli éclater de rire quand celui-ci l'a déclamé à son tour. Je pensais : « C'est Gide qui est là, qui me parle, qui raconte des souvenirs de La Roque » (qui a appartenu à la famille de Racine) ; j'avais peine à m'en persuader, tant je trouvais cette rencontre naturelle et familière cette conversation.*
>
> *Et de même lorsqu'un peu plus tard il nous a appelés dans sa chambre pour nous lire les fragments de Hugo qu'il n'a pas conservés pour son anthologie et qu'il appelait les « tombées de Dieu » (« Je vais vous lire les tombées de Dieu ») - et même en voyant ces vers écrits de la célèbre écriture, j'arrivais mal à me persuader que j'avais devant moi l'homme qui avait joué un si grand rôle pour ma jeunesse. Tout devenait intime, aisé, comme connu depuis longtemps."*

English translation:

> *"He came back a little later, just as Thomas was reading to me the passages from The End of Satan he had told me about during the journey, and in particular the famous verse:*
>
> *Whale with an infected and dilapidated jaw.... [God, II, 7]*
>
> *which he claimed by imitating Gide - and I almost burst out laughing when Gide in turn claimed it. I thought: "It is Gide who is there, who speaks to me, who tells me about memories of La Roque" (who belonged to Racine's family); I had difficulty to convince myself of this, so much I found this meeting natural and familiar this conversation.*
>
> *And the same when he called us into his room a little later to read us the fragments of Hugo that he did not keep for his anthology and that he called the "fallen of God" ("I will read to you the fallen of God") - and even when I saw these written verses from the famous writing, I could hardly convince myself that I had before me the man who had played such a great role for my youth. Everything became intimate, easy, as it had been known for a long time."*



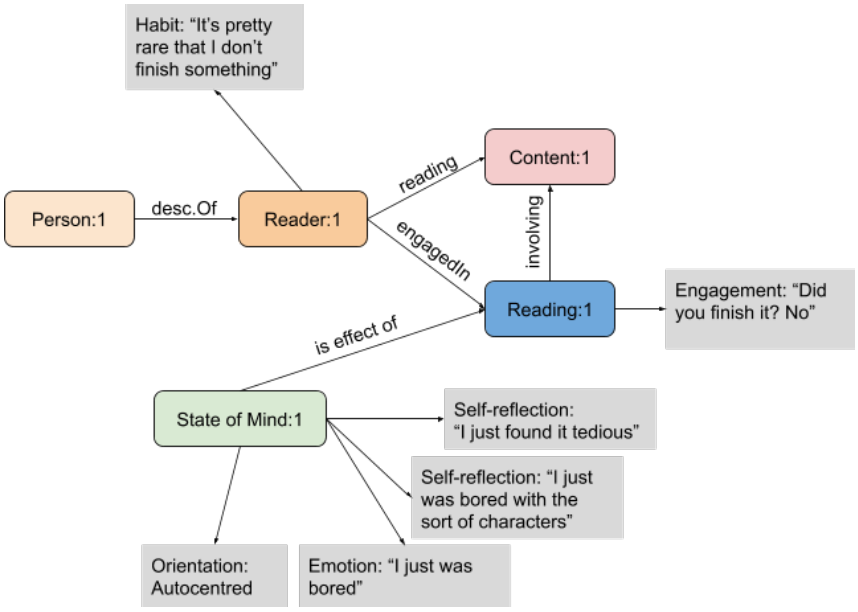
Simplified graph representation of the extract (State of Mind annotated but omitted). The features of Entities (e.g. Transportation) are fragments extracted from the source (e.g. "I just was swallowed up by it") and structured values represented with grey boxes (e.g. Location: Italy).

Non-readings, attempts to read and brief readings

In most of the cases, we as reader do not go fully through the reading we start, and actually we have very brief interactions with a content. This form of interaction is common when we perform an assessment of a content, such as in case we visit a book shop or library, or in research when you evaluate the relevance of papers to consider. Furthermore, there are cases in which we wish to engage in a reading, due to the lack of interest or opportunity. It is worth to mention that the title and a summary of the book are respectively part of the work and part of the expression (in the frame of FRBR conceptualisation). Thus, these scenarios are indeed readings too and should be encoded as such.

Source : Memories of fiction, Johanna W. interview part 1 [transcribed], 2015, https://soundcloud.com/memoriesoffiction/johanna-williams-1

> *"Are there any particular genres that you enjoy or any that you really don't like?*
>
> *I didn't like North West, I just, by Zadie, can't remember her surname, I just found it tedious.*
>
> *Really? Why was that?*
>
> *And there's very few books [coughs], I don't know, but I just was bored with the sort of characters, really, just …*
>
> *Did you finish it?*
>
> *No, which is …*
>
> *So, if you don't like something, you don't finish it?*
>
> *It's pretty rare that I don't finish something, but I just thought, I'm wasting my time reading this.*
>
> *Yeah, 'cause I think it was, that question of whether to plod on or fling it, you know [laughter] …*
>
> *I know, and then I think, yeah, and then I just thought, oh, life's too short to read this, and that was a book club book."*



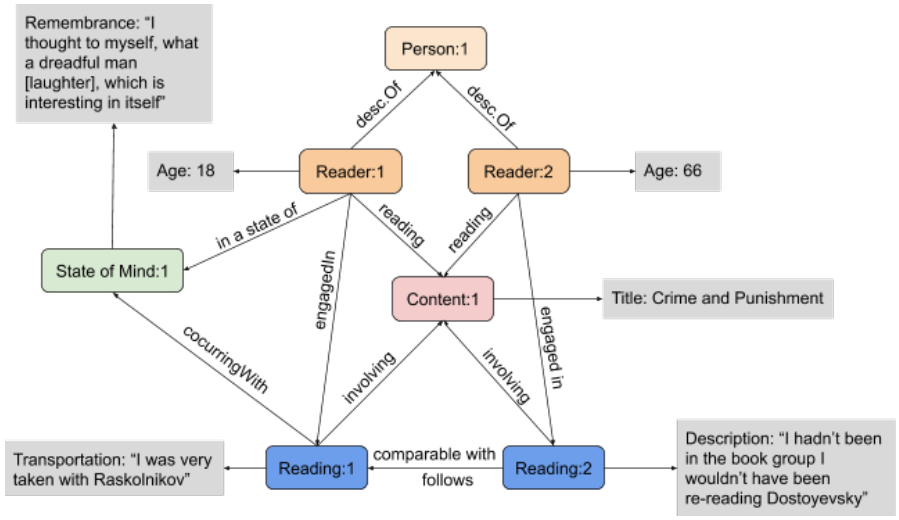
Simplified graph representation of the extract. The features of Entities (e.g. Transportation) are fragments extracted from the source (e.g. "I just was swallowed up by it") and structured values represented with grey boxes (e.g. Location: Italy).

Comparative readings

A reading experience is interweaved with previous reading, we compare memories and outcomes highlighting the features by difference. This form of reporting a reading experience is common of reviews, e.g. comparing different works of the author and serial novels. This scenario can be encoded by considering *Readings* or *Reading Sessions* as *Events* and representing the relations such as *comparable with, better than* and *similar to*.

Extract from Memories of fiction, Brenda C. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/brenda-part-2.

> *"I was reading the Dostoyevsky again, I read it in the sixth form, and this is another aside, but I was very taken with Raskolnikov, you see, when I was eighteen. I thought he was a bit of a, you know. I read it again when I was sixty-six, and I thought to myself, what a dreadful man [laughter], which is interesting in itself, isn't it, how you have different views, and how badly exploited those women were [laughter], do you see, I hadn't even thought of the role of women when I was a teenager, I just saw this poverty stricken, wonderful student with this in his head, and how he's been brought low by his poverty. I mean, you know, I mean that's a crude interpretation of what I thought, but I think, you know, but I was very much, so that was interesting, but anyway, I was reading that and then didn't get to discuss it sadly, but it was, you know, it's about redemption, isn't it, and I think, and, but at the same time I was coincidentally reading Donna Tartt's latest, The Goldfinch, and I could say the frame of Dostoyevsky in that novel. That was about redemption, wasn't it, and the symbol of that bird, which, you know, the goldfinch, tell me, I don't want to sound like, you know, I'm telling you something, stating the obvious, but I'm saying because in the painting the goldfinch is Christ, is used as Christ, isn't he, with the sash of blood across his breast, and that, you know, choosing that as the metaphor for the novel, and that painting as the metaphor, it was about that boy's redemption, wasn't it, his suffering and redemption through suffering, and that is what happened to him, that's what happened to Raskolnikov, wasn't it, so I saw these immediate parallels between those two novels, which was interesting because if I hadn't been in the book group I wouldn't have been re-reading Dostoyevsky."*



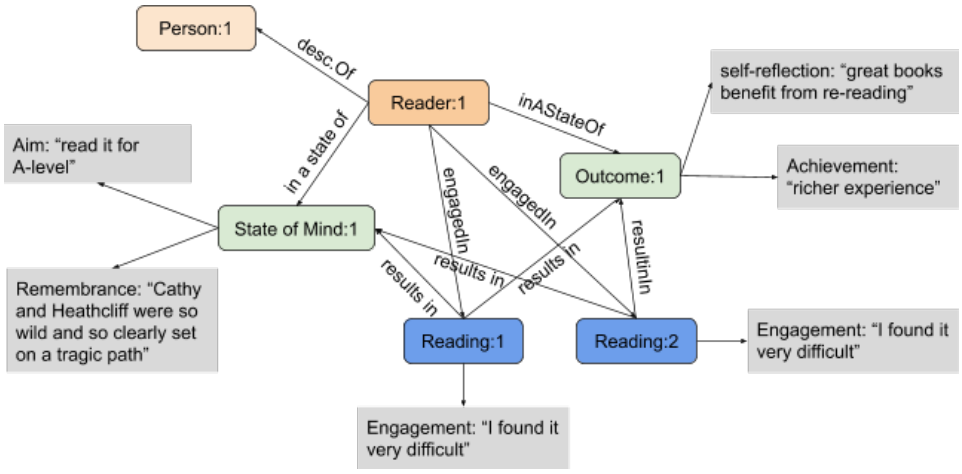
Simplified graph representation of the extract. The features of Entities (e.g. Transportation) are fragments extracted from the source (e.g. "I just was swallowed up by it") and structured values represented with grey boxes (e.g. Location: Italy).

Cumulative outcomes of sequential readings

There are contents we read multiple times, in different moment of our life, leading us toward a cumulative experience as synthesis of these multiple engagements. This type of experience is indeed similar to "comparative readings", but in this case the accent is not the comparison between readings while a set of Reading(s) *resulting in* a collective Outcome.

Extract from Memories of fiction, Ferelith H. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/ferelith-part-2.

> *"[It] took me a long time to read Wuthering Heights.*
>
> *Because you found it difficult?*
>
> *I found it very difficult. Cathy and Heathcliff were so wild and so clearly set on a tragic path, and it wasn't 'til I read it for A-level and then had to re-read it and re-read it and re-read it, and it is a book that really re-pays re-reading, and every re-reading made it a richer experience. It was one of the lessons that one learnt from A-level that actually great books benefit from re-reading. Re-reading can be a really, really good thing. Having said that I'd always re-read. I, I used to re-read Heidi, which, really the most dreadful book, but I used to consciously say, do I wish to cry, or do I wish to not cry. If I felt like crying, I would read her being sent to Frankfurt. If I didn't want to cry I would skip that chapter."*



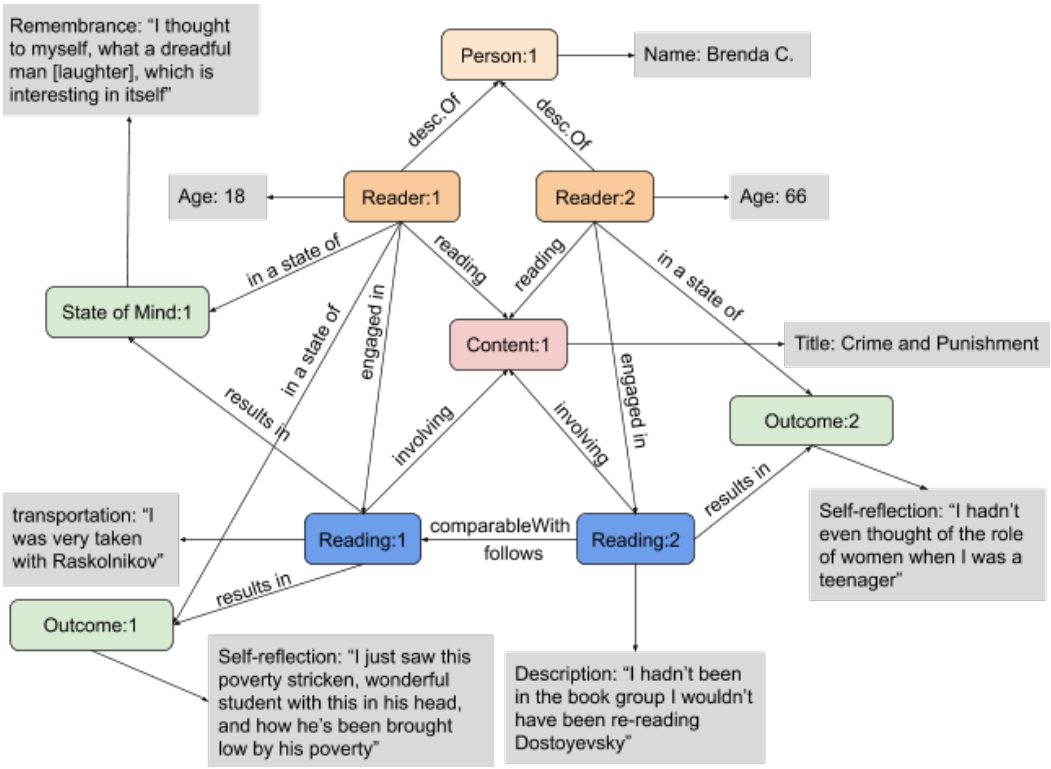
Simplified graph representation of the extract. The features of Entities (e.g. Transportation) are fragments extracted from the source (e.g. "I just was swallowed up by it") and structured values represented with grey boxes (e.g. Location: Italy).

First and last impressions of reading

As result of reading we develop a first impression which is likely to be replaced by a last impression after a certain amount of time and self-reflection. Interviews and reviews often report elements of the first impression and of the last impressions, highlighting the evolution of the reader's evaluation and how the weight in these evaluations switch from a set of elements to others. This type of evidence is a kind of comparative evaluation between multiple *Outcomes* at different time of a single *Reading*.

Extract from Memories of fiction, Brenda C. interview part 2 [transcribed], 2014, https://soundcloud.com/memoriesoffiction/brenda-part-2.

> *"I was reading the Dostoyevsky again, I read it in the sixth form, and this is another aside, but I was very taken with Raskolnikov, you see, when I was eighteen. I thought he was a bit of a, you know. I read it again when I was sixty-six, and I thought to myself, what a dreadful man [laughter], which is interesting in itself, isn't it, how you have different views, and how badly exploited those women were [laughter], do you see, I hadn't even thought of the role of women when I was a teenager, I just saw this poverty stricken, wonderful student with this in his head, and how he's been brought low by his poverty. ..."*



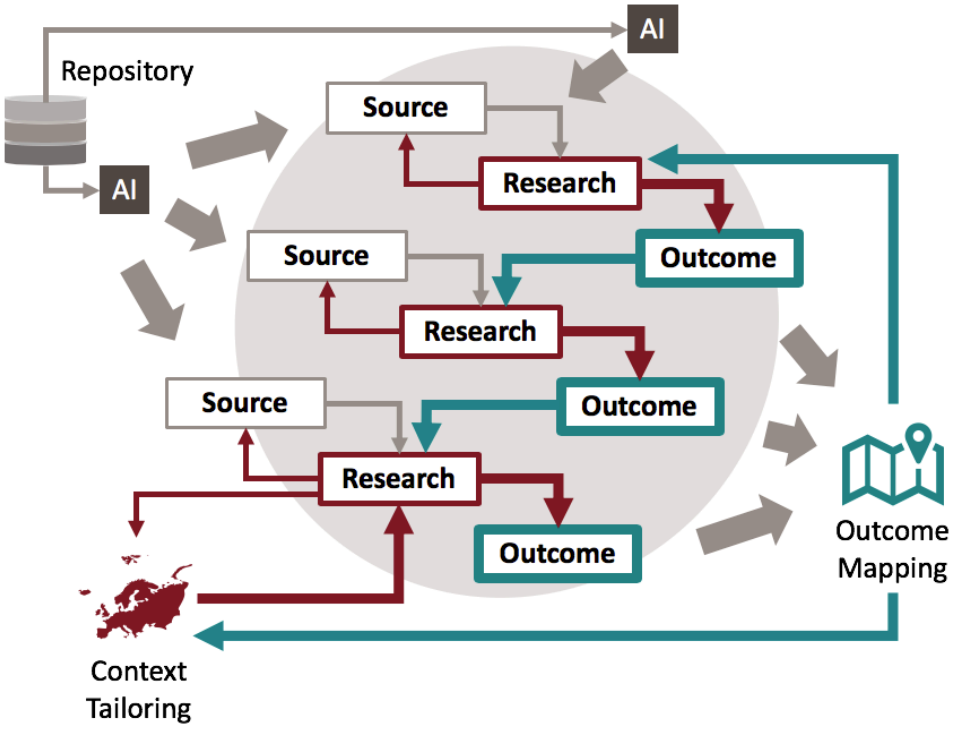
Simplified graph representation of the extract. The features of Entities (e.g. Transportation) are fragments extracted from the source (e.g. "I just was swallowed up by it") and structured values represented with grey boxes (e.g. Location: Italy).

Conclusions and Future Developments

READ-IT data model has the ambitious aim to enable the creation of a database of European reading experience as benefit of future researches and applications. Among the many issues we must face in the design and development of READ-IT database, the definition of a data model for the description of such a complex phenomenon, reading, is among the most challenging. Indeed, the model of reading must 1) define a phenomenon that is still object of study, 2) make a synthesis of orthogonal perspectives on reading and 3) it cannot rely on a consolidated theory of reading, but must take into account an impressive literature about specific aspects of reading from multiple disciplines.

In this frame, we adopted a methodology based on the analysis of sources to guide the identification of the concepts of the model of reading. Following, we used theories from multiple disciplines to guide the integration of these concepts and, lastly, we stressed the model through the direct engagement of researchers in READ-IT. The resulting model is not final, but a baseline to support the development of the case studies, which outcomes will probably drive the revision of the model at late stage of READ-IT.

At the current stage of the project, we already foresee new lines of research related to READ-IT database. Firstly, the development of tools for taming the data to support the retrieval (recommender systems) and the generation of data from new sources. Specifically, this second strand of work should aim to the development of tolls for supporting the annotation of new sources, assisted or automatic annotation, by exploiting machine learning algorithms and the data collected in READ-IT database.



The future developments of READ-IT could follow the use of AI techniques for automatic annotation of sources, and recommendation of sources. Furthermore, it is worth to consider extending the scope of the database to research outcome and provide tools for the exploration (outcome mapping and context tailoring) of reading research field.

The second line of research should be focused on technologies for navigate the complexity. Specifically, the exploration of research outcomes aiming to support researchers in identifying the status of the research on reading and the gaps that must be addressed, e.g. in terms of type of sources, periods, locations, facets of reading. Furthermore, these tools for exploration should be able to provide a context tailored on the framing of new case studies based on the reuse of sources and outcomes.

Annexes

Annex 1 - Question & Answers

Is the reader's writing habit within the scope of the model?

No, it is not. Reader's *habits* or patterns in reporting their Experience is not an evidence, but something that could be inferred by a set of evidences. Since it is an inference, it should be the result of querying data. If the question is if it will be possible to infer such habits from the data, the answer is yes in principle.

The reader is defining his/her experience by comparison. Does the model support comparison between experiences?

Yes, we introduced a *reference* relation between Reading(s), Sessions and Experience(s). The reference could be generic or further specified, e.g. *better than, referred by*.

In case the evidences of reading are meant for someone, e.g. the author, how this information could be represented?

A reading could be framed in an activity such as a book review or correspondence. This information should be encoded as aim, activity (as Premise of the Reading) and *achievement judgment* or *self-reflection* (as Outcome of the Reading), considering what the evidence reports.

In the evidence, the reader is describing emotional aspects of the content and not of his/her personal reading experience. Is this information in the scope of READ-IT data model?

In the model there are yet no way to separate a description of the content from the reader's personal experience. This is related to the assumption that a description is in fact a reader's interpretation of the content, therefore not objective but informative of the experience.

What is the minimum set of information required to have a reading experience entry?

The minimum unit is the evidence of a reading activity, i.e. an actor / content interaction. Any source reporting directly or indirectly a reading is valid.

How experimental data should be encoded in the model?

Experimental data are a source of evidence of reading. Therefore, experimental data should be encoded as target in the W3C Web Annotation Data Model. If the experimental data had been annotated manually or through an algorithm, the W3C Web Annotation Data Model offers the concept of agent to report the person(s) or the algorithm(s).

My evidence reports of reading social media, blog posts or other contents which are not books, how should that be encoded in the data model?

Everything read should be encoded as content. In READ-IT model, content is an extension of FRBR work, expression and manifestation, but to extend the support to other sources the content is also an extension of Schema.org creativeWork, the superclass of Article, BlogPost, Question, Review, etc.

How should I specify if a reader is an expert and/or if the reader is reading as part of his/her professional activities?

The expertise can be encoded as expertise of a Reader or as Reading part of a professional activity. In the first case, the expertise (or skills) of the reader can be encoded by using the features of the Reader, e.g. occupation, education. In the second case, the Reading can be characterised by a Premise (State of Mind) defining the aims related to the professional activities of the reader. Furthermore, the professional activities of a reader, e.g. a project or exam, can be encoded using the concept of Event *involving* (or *co-occurring with*) a Reading.

Is there a distinction between fiction and non-fiction content?

there are no a priori differences between in reading fiction and non-fiction content. It could have a different effect on the reader, but as result of the motivations for reading and the all other factors affecting the evaluation of a reading.

How can I quantify the reader's engagement?

The engagement on the reading can be encoded using the properties *engagement* and *transportation* of the Reading, Session and Experience. With *engagement* we intend the general participation of the Reader in the Reading or Session as activities, while with *transportation* we refer to the engagement of the Reader in the content (e.g. narrative).

Annex 2 - References

- Adamou, A et al., (2018). "Crowdsourcing Linked Data on listening experiences through reuse and enhancement of library data." *International Journal on Digital Libraries* (2018): 1-19.
- Hitchcock, T. (2014). 'Big Data, Small Data and Meaning. http://historyonics.blogspot.co.uk/2014_11_01_archive.html. 9 November 2014.
- Kovecses, Z. (2003). *Metaphor and emotion*. Cambridge UP
- Lakoff, G., Thomson, M., 2003. *Metaphors we live by.* University of Chicago Press
- Towheed, Shafquat (2011). 'Introduction'. In *The History of Reading, Volume 3: Methods, Strategies, Tactics*, edited by Rosalind Crone and Shafquat Towheed, 1–12. Basingstoke, UK: Palgrave Macmillan.
- Craig, Robert P. (1984). Developing a Philosophy of Reading: Piaget and Chomsky. Reading Horizons, vol. 25, issue 1.
- Norman, D. (1986). "User Centered System Design: New Perspectives on Human-computer Interaction". CRC. ISBN 978-0-89859-872-8.
- Evans, C. David (2017). Bottlenecks - Aligning UX Design with User Psychology, pages 80-81. Apress, DOI: 10.1007/978-1-4842-2580-6.
- Hutchins, E. (2009). Cognitive Ecology. Topics in Cognitive Science, vol. 2 (2010), 705-715. DOI: 10.1111/j.1756-8765.2010.01089.x.
- Golonka, S., Wilson, A. D. (2012). Gibson's ecological approach = a model for the benefits of a theory driven psychology. AVANT, vol. III, no. 2/2012.
- Dourish, P. (2001). Where the action is. Cambridge: MIT press.
- Gibson, J. J. (1979). The Ecological Approach to Visual Perception. Houghton Mifflin Harcourt (HMH), Boston.
- Forlizzi, J. (2008). The Product Ecology: Understanding Social Product Use and Supporting Design Culture. International Journal of Design, 2(1), 11-20.
- Ferraris, M. (2013). Documentality. Fordham UP.
- Gerrig, R. J. (1993). Experiencing Narrative Worlds - On the Psychological Activities of Reading. Yale University Press.
- Gibson, W. (1980). Authors, speakers, readers, and mock readers. In J.P. Tompkins (ed.), Reader-response criticism (pp 1-6). Baltimore: Johns Hopkins University Press.
- Stanislavski, C. (1936). An actor prepares. New York: Theatre Arts.
- Ferraris, M. (2008). Science of Recording. Publications of the Austrian Ludwig Wittgenstein Society - New Series Vol. 7.
- Austin, J. L. 1962. How to do things with words. London: Oxford University Press.
- Searle, J. 1995, The Construction of Social Reality, New York, Free Press.
- Davies, D. (2007). Aesthetics and literature. Continuum International Publishing Group.
- Bortolussi, M., & Dixon, P. (2007). Psychonarratology. Cambridge University Press.
- Calvino, I. (1979). Se una notte d'inverno un viaggiatore. Einaudi.
- Green, M., Strange, Jeffrey J, & Brock, Timothy C. (2002). Narrative impact : Social and cognitive foundations. Mahwah, NJ: L. Erlbaum Associates.
- Pennac, D. (1992). Comme un roman. éditions Gallimard.
- Estetica (2009). Dizionario di filosofia. Treccani. Visited on December 2018, URL: http://www.treccani.it/enciclopedia/estetica_%28Dizionario-di-filosofia%29/.
- Etica n.d. Enciclopedie on line Treccani. Visited on December 2018, URL: http://www.treccani.it/enciclopedia/etica/.
- Hartmann, N. (1970). Etica: Assiologia dei costumi, a cura di Vincenzo Filippone-Thaulero, Guida, Napoli.
- Bara, B. G. (1999). Pragmatica cognitiva. Bollati Boringhieri.
- IFLA (2009). Functional Requirements for Bibliographic Records - Final Report. International Federation of Library Association and Institutions, url: http://www.ifla.org/VII/s13/frbr/, visited November 2018.
- Wittgenstein, L. (1953). Philosophical Investigations.

- Rosch, E. (1983). Prototype Classification and Logical Classification: The Two Systems.,in Scholnick, E.K. (ed), New Trends in Conceptual Representation: Challenges to Piaget's Theory?, pp. 73–86, Lawrence Erlbaum Associates, Hillsdale.
- Grice, H. P. (1989) Studies in the Way of Words. Cambridge (MA): Harvard University Press.
- Marrone, G. (2001) Corpi Sociali: processi comunicativi e semiotica del testo. Einaudi.
- Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. Cognitive psychology, 43(1), 1-22.
- Driscoll, B., & Rehberg Sedo, D. (2018). Faraway, So Close: Seeing the Intimacy in Goodreads Reviews. Qualitative Inquiry, 1077800418801375.
- Eco, U. (2011). Lector in fabula: la cooperazione interpretativa nei testi narrativi. Giunti.
- Davis, L. H. (1979). Theory of Action. Prentice-Hall, Inc.
- Gibson, J. J. (1979) The Ecological Approach to Visual Perception. Houghton Mifflin.

Annex 3 – Changelog

16/11/2018	Alessio Antonini	Creation of the document, general structure, Q&A
26/11/2018	Alessio Antonini	Glossary, modelling principles and description of the model
27/11/2018	Alessio Antonini, François Vignale	Introduction
28/11/2018	Alessio Antonini, Guillaume Gravier	Web Annotation Data Model
28/11/2018	Alessio Antonini	Standards & Design Patterns draft, Causal theory of reading, Extensional & Intensional description of reading, Reading Interaction and Experience
12/12/2018	Alessio Antonini	Data model graph, Reading Process, recording of readings
12/12/2018	Alessio Antonini, François Vignale	Introduction of Anatomy of Reading, proofreading
17/12/2018	Alessio Antonini, Fraçois Vignale	Proofreading
17/12/2018	Alessio Antonini	Q&A, Standards and Design Patterns, Conclusions
18/12/2018	Alessio Antonini, Fraçois Vignale	Proofreading
18/12/2018	Alessio Antonini	Executive summary
20/12/2018	Alessio Antonini, Fraçois Vignale, Guillaume Gravier	Revision of annotation model, introduction of Reading Interaction, skeleton of section Use Patterns
08/01/2019	Alessio Antonini	Change document structure
10/01/2019	Alessio Antonini	Fixing of annotation description. Schema.org event. Description of Design Patterns. New subsection about Accessing to Contents. Introduction to the epistemology. Introduction of the anti-design patterns (misuses) as follow-up of the epistemology. Fix Q&A about reader's engagement.
15/01/2019	Alessio Antonini	Removed section about Research with READ-IT data and Anti-Patterns. Swap of order between epistemology and standards.
27/01/2019	François Vignale, Alessio Antonini	Proofreading
28/01/2019	François Vignale	Data model and design pattern examples
05/02/2019	François Vignale	Data model and design pattern examples
12/02/2019	Alessio Antonini	New graphs in Data Model
13/02/2019	Alessio Antonini	New structure and graphs in Data Model. Alignment with FRBR, Schema.org and FoaF.
14/02/2019	Alessio Antonini	Fix Executive Summary. Finalized the Action of Reading in The Data Model.

15/02/2019	Alessio Antonini	Finalize the Process of Reading in the Data Model section.
26/02/2019	Alessio Antonini	Fix Example 1, annotation of design patterns, removed epistemology section
12/03/2019	Alessio Antonini	Example 2, annotation and graph of Multiple readers, Non-readings
13/03/2019	Alessio Antonini	Layout and proofreading
16/03/2019	Brigitte Ouvri Vial	Proofreading
06/05/2019	François Vignale	Proofreading
15/05/2019	François Vignale	Fix The Reader as Agent
16/05/2019	François Vignale	Proofreading. Fix Temporal entities
08/07/2019	François Vignale, Brigitte Ouvry-Vial	Proofreading.
09/07/2019	Alessio Antonini	New structure
10/07/2019	Alessio Antonini	New structure of annex, extending and fixing the introduction, notes for the conclusions, CIDOC-CRM description and alignment
11/07/2019	Alessio Antonini	Introduction of the model of reading about the adopted approach
12/07/2019	Alessio Antonini	Rewriting of the Model of Reading section: new structure, examples, diagrams and description of concepts.
15/07/2019	Alessio Antonini	Reading habit, engagement in reading, introduction to the data model, remove listening from examples, new summary
16/07/2019	Alessio Antonini, Fraçois Vignale	Add orientation of the State of Mind
16/07/2019	Alessio Antonini	Fix examples, fix intro data model, fix Q&A
17/07/2019	Alessio Antonin, François Vignalei	Proofread, Release of the document for internal use