9^{th}ArtOnto: Towards an ontology for representing visual and narrative structures

Alexandre Jaud[1], Jean-Christophe Burie[1], Vincent Nguyen[2], Christophe Rigaud[1,3], and Samuel Petit[1,3]

[1] L3i Laboratory, SAIL joint Laboratory, 17042 La Rochelle CEDEX 1, France
[2] University of Orléans, LIFO UR 4022, INSA-CVL, Orléans, France
[3] Mangas.io, 14000 Caen, France
alexandre.jaud@univ-lr.fr

Abstract. We present 9^{th}ArtOnto, an ontology that provides a formal framework for the construction of knowledge graphs representing comic book works. The ontology captures both the visual and structural content of a work, including pages, panels, characters, objects, balloons, captions, and onomatopoeia, and encodes the narrative layer through the grouping of panels into coherent scenes and narrative arcs. This dual representation enables the reconstruction of both the *fabula* (the chronological order of events in the story world) and the *syuzhet* (the order in which these events are presented to the reader), including non-linear structures such as flashbacks. We present an initial demonstration of the ontology's ability to jointly represent the visual and narrative dimensions of a sequential art work within a single, queryable knowledge graph. The resulting structured representation opens the way to several downstream applications, including context-aware accessibility tools for visually impaired readers, automated summarization, cross-work narrative comparison, and the grounding of large language models in the structured content of comic book works.

Keywords: comics understanding · narrative · ontology · knowledge graph.

1 Introduction

Sequential art, also known as the 9^{th} art in France, is a complex medium encompassing manga, bandes dessinées, webtoon, manhwa, graphic novels, and comics, in which textual and visual information are organized according to specific narrative conventions. Over the past years, a growing body of research has focused on the detection of individual comic book elements, such as panels, speech bubbles, and characters [26, 27]. However, comparatively little attention has been paid to the structural relationships between these elements and their connection to the underlying narrative.

Establishing such a structure would make it possible to enumerate the complete set of visual information present in a work, thereby enabling downstream

applications that require access to the full or partial context of a comic book. Knowledge graphs provide an effective means of representing information in a structured and semantic manner [18]. However, a robust knowledge graph requires an ontology that captures the relevant concepts and relationships within a domain, serving as the formal backbone of the knowledge model [17]. To address this gap, we propose an ontology designed to represent both the visual and narrative dimensions of sequential art within a unified and structured framework, despite the inherent complexity of this medium and the artistic freedom exercised by authors, who often deliberately subvert established conventions.

Our contributions are as follows:

- We introduce $9^{th}ArtOnto$, an ontology providing a formal framework for the construction of a knowledge graph that captures the narrative and visual dimensions of a sequential art work.
- We present a preliminary implementation of this framework applied to a single manually annotated comic book, offering an initial illustration of how both dimensions can be represented within a single knowledge graph.

The remainder of this paper is organized as follows. Section 2 reviews related work on graph-based approaches in comics, existing annotation structures, and narrative ontologies. Section 3 presents the ontology used to structure the knowledge graph. Section 4 describes the implementation of this ontology to a comic book. Section 5 discusses the limitations of the proposed ontology. Finally, Section 6 outlines potential applications and future directions for extending the proposed ontology.

2 Related works

2.1 Narrative Ontologies

Several ontologies have been developed to formally represent narrative structures, primarily in the context of computational literary studies and digital humanities. Among the most recent is GOLEM [21], which provides a structured and computationally tractable representation of narrative and fictional elements. Built upon CIDOC CRM [11], LRMoo [1], and the DOLCE [13] upper-level ontology, it captures narrative structure, character dynamics, and fictional worlds, while supporting provenance tracking and pluralistic interpretations. Its modular design facilitates alignment with various literary and narratological theories, making it a flexible framework for comparative analysis across texts and media; its *Social Relationship* and *Relationship Role* modules in particular address dimensions that fall outside our current scope but represent natural directions for future extensions (see Section 5). The Narrative Ontology (NOnt) [20] formalises the classical narratological distinction between *fabula*, *narration*, and *reference* at a level of abstraction well suited to digital libraries, while ontologies such as the Drammar ontology [10] address more domain-specific narrative structures. However, none of these frameworks provides any mechanism for grounding narrative

constituents in visually annotated content, nor for representing the page-level and panel-level structure specific to sequential art. They are therefore ill-suited to serve as the basis for constructing a knowledge graph directly from a comic book annotation, which is the primary objective of *9^{th}ArtOnto*.

2.2 Comic Ontologies

An early ontology-based approach to comic book image analysis is proposed by Guérin et al. [15, 16], who organize their framework around two complementary ontologies: a comics-specific ontology and a generic image ontology. The former introduces a set of core classes, namely *Comic*, *Plate*, *Panel*, *Character*, *Balloon*, *TextLine*, and *Tail*, while the latter formalizes fundamental image processing concepts, such as images and regions of interest, in a domain-agnostic manner. These two ontologies are connected through formal equivalence relations, enabling iterative content discovery within an image analysis pipeline. Although this work constitutes a solid foundation for the structural description of comic book pages, its narrative dimension remains partial: it does not address the modeling of events, inter-character relationships, or narrative arcs, as the primary objective is image analysis and retrieval rather than narrative understanding. Furthermore, the framework is designed around conventional comic book layouts and does not account for more complex elements such as multi-layered speech balloons or onomatopoeia.

Chen [5] adopts a more narratively oriented perspective, proposing a hierarchical knowledge graph framework for the structured semantic understanding of visual narratives. This architecture organizes narrative content across three distinct levels of abstraction. At the panel level, nodes encode both visual entities, such as characters and backgrounds, and textual components like dialogues and captions. These elements are then linked at the sequence level through a temporal graph; by utilizing a directed acyclic graph with *precedes* relations, the framework effectively represents non-linear structures like flashbacks. Finally, at the event level, Chen's framework organizes the narrative into a three-tier hierarchy grounded in Cohn's Visual Narrative Grammar [8], spanning from macro-events down to fine-grained segments. Cross-level relations, including *instantiates* and *subevent-of*, integrate these tiers into a single, unified representation.

2.3 Comic Encoding

A notable effort in the structured encoding of comics is the Comic Book Markup Language (CBML) [28], an XML vocabulary extending the TEI (Text Encoding Initiative) P5 guidelines to encode comic book content in a document-centric format. CBML introduces a small set of dedicated elements, principally `cbml:panel`, `cbml:balloon`, and `cbml:caption`, allowing the encoding of the physical and textual components of a page. The hierarchical structure of the document is implicitly carried by the XML nesting, which reflects the containment relationships between these elements. This vocabulary remains limited in scope; it captures the surface layout of comics pages but does not provide any formal

modeling of the relationships between characters, the reading order as an explicit structure, or the narrative dimensions of the work. As a transcription-oriented format, CBML is well suited to the digitization and archival of comic books in the tradition of digital humanities, but it does not constitute a knowledge representation framework and cannot serve as the basis for structured reasoning over the content of a work.

3 Proposed ontology: 9thArtOnto

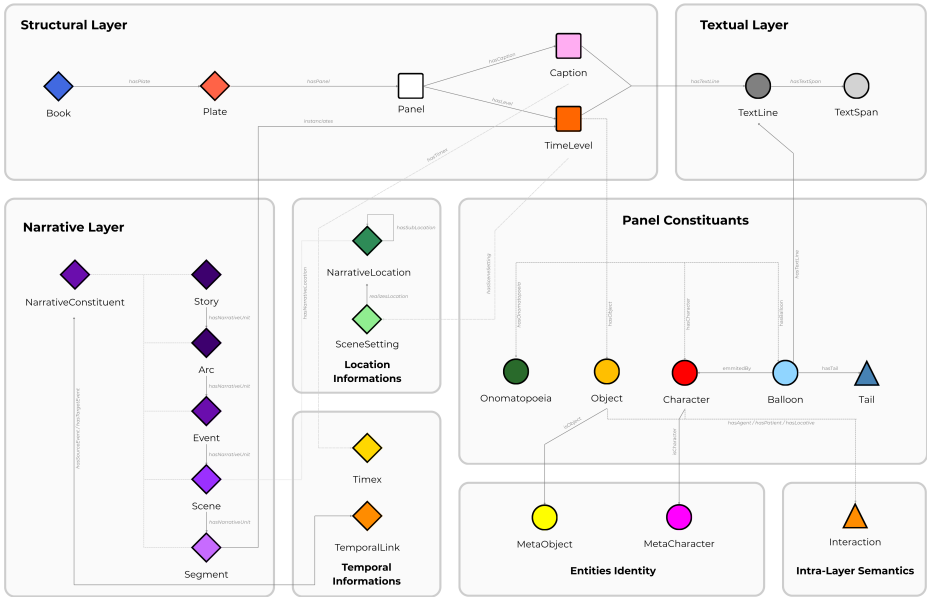

Fig. 1. Overview of the general principles of *9th ArtOnto*.

We present *9th ArtOnto*, an OWL 2 [14] ontology dedicated to the joint representation of the visual, structural, and narrative dimensions of sequential art works. OWL 2 is a W3C standard for representing knowledge as formal ontologies, grounded in description logics, which enables automated reasoning over structured data. The ontology is serialised in RDF/XML and edited with Protégé[4]. At the logical level, the underlying RDF layer represents every piece of knowledge as a *(subject, predicate, object)* triple, while OWL 2, grounded in description logics, provides the formal machinery needed to declare classes, object properties, datatype properties, and the constraints that relate them. *9th ArtOnto* is deliberately designed as an application ontology whose primary objective is the construction of queryable knowledge graphs over comic book works. Formal alignment with upper-level ontologies such as CIDOC-CRM or DOLCE, conceived for interoperability in digital humanities infrastructures and library

[4] `https://protege.stanford.edu/` (accessed June 2026)

systems, falls outside this scope. An overview of the main components and general principles of the ontology is shown in Fig.1.

3.1 The visual and structural layer

The visual and structural layer is the largest part of the ontology. We build upon the comic-specific ontology of Guérin et al. [15, 16], which provides a solid backbone for page-level description, and we extend it to accommodate artistic conventions that are poorly handled by existing models such as multi-layered balloons, onomatopoeia, borderless panels, and panels that encode several temporal strata at once.

From Book to TimeLevel. The top-level hierarchy runs from the work as a whole down to the spatio-temporal strata that group graphical constituents together. The `Book` class denotes the work as a whole and is deliberately renamed from Guérin's `Comic` to a neutral umbrella term covering manga, bandes dessinées, and comics. We retain a `title` datatype property and a `right2Left` boolean that carries reading direction (crucial to distinguish manga from Western comics), and introduce a `graphicStyle` property that characterises the overall visual register of the work (*e.g.* black and white, colour, watercolour, realistic, impressionist). Bibliographic metadata (author, ISBN, publisher) are deliberately excluded from the core model as they are not relevant for the spatial and narrative analysis we target, but may be added as extensions. A book is linked to its pages through the `hasPlate` object property.

A `Plate` represents an individual page. Beyond the classical attributes `hasNumber` and `onDoublePage`, we introduce an `atmosphericLayer` property capturing the general ambience perceived when discovering the page at first glance (*e.g.* an entire plate covered in snow, a red-tinted page signalling danger). This property is currently free-form but is intended to be later constrained by a controlled taxonomy. Pages are chained via `hasNextPlate` and linked to their panels via `hasPanel`.

The `Panel` class is the pivotal entity of the ontology. Rather than a rectangular bounding box, each panel is described by a *mask*, since panels frequently adopt non-rectangular shapes (*e.g.* diagonal borders in manga). In addition to the ordering properties `hasRank` and `hasNextPanel` inherited from [16], we introduce a `borderType` attribute to encode whether panels are fully bordered, partially bordered, or borderless, and a panel-level `graphicStyle` which, when non-null, signals a local deviation from the book-level style (*e.g.* a sepia shift for a flashback). We further extend the model along axes drawn from comics theory: transitions inspired by McCloud [25] and panel functions from Cohn [6, 9], giving rise to panel-transition type and four specialised panel types:

- **Transitions.** McCloud's six-fold typology (action-to-action, aspect-to-aspect, moment-to-moment, non-sequitur, scene-to-scene, subject-to-subject) is encoded as a `transitionType` attribute relative to the previous panel.

- **Polyptych panels.** A polyptych fragments a unique and continuous decor across several panels, while one or more characters appear multiple times at different instants within that shared background. We encode this with a boolean `isPolyptic` and an object property `hasPolypticContinuity` linking the panels belonging to the same polyptych.
- **Locative panels.** Panels whose primary function is to establish the spatial context of a scene without advancing the action are flagged as *locative*.
- **Polymorphic panels.** A polymorphic panel depicts the same character at several successive instants within a single panel, superimposing or juxtaposing different states of one action.
- **Polychronic panels.** We use the term *polychronic* to designate panels in which at least two diegetically separate time-points are simultaneously rendered within the same panel. Unlike polymorphic panels, which superimpose successive states of a single character's motion, a polychronic panel does not necessarily repeat the figure: its defining feature is the diegetic nature of the temporal split. A foreground character may occupy the narrative present while the background simultaneously depicts a past event, embedding a flashback within the confines of a single encapsulated image rather than delegating it to a dedicated panel [4]. An annotated example is presented in Section 4.

The existence of polychronic panels breaks the usual one-to-one correspondence between a panel and a single spatio-temporal situation, making it necessary to group graphical constituents into internally coherent strata rather than attaching them directly to the panel. We introduce the `TimeLevel` class for this purpose. Unlike a mere spatial layer, a `TimeLevel` is defined by a chronological and diegetic function, and a panel is linked to its levels via `hasTimeLevel`. In the overwhelming majority of cases a panel has exactly one `TimeLevel`, in which space and time are unified; polychronic panels are the exception that requires several. In what follows, every graphical constituent of a panel is attached to at least one `TimeLevel` rather than directly to the panel itself.

3.2 The graphical constituents of comics

Within each `TimeLevel`, the graphical constituents are organised around **agentive** and **non-agentive entities**, together with the speech and sound elements they produce or carry. Characters in sequential art are not necessarily human-like or animal-like: robots and anthropomorphised objects can all qualify [3], provided they are ***agentive*** (*i.e.* intentions, beliefs, or autonomous decisions can be attributed to them) and carry a narrative function. To ensure identity persistence across panels and pages, we adopt a two-level pattern: a `MetaCharacter` node represents the identity of the character as it persists throughout the whole work (*e.g.* a unique MetaCharacter for *Asterix*, *Naruto*, or *Superman*), while a `Character` node represents a local occurrence of that identity within a specific `TimeLevel` and panel; successive occurrences may exhibit variations in appearance, posture, expression, or drawing scale due to the artist's stylistic

freedom. Every `Character` instance is linked to its `MetaCharacter` through an `instantiates` property. Crucially, a `Character` carries a `visible` boolean: setting it to *false* allows the modelling of off-panel entities that are diegetically present without being graphically depicted, such as a speaker whose balloon appears while the character lies outside the frame.

To push beyond a purely structural description of the page, we include ***non-agentive entities*** through the `Object` class. Modelling every depicted object indiscriminately would flood the graph with irrelevant entities; we therefore restrict `Object` to entities that either (i) directly interact with a `Character` in the panel, or (ii) have an explicit narrative role: named, recurrent, triggering an event, or carrying symbolic meaning. Persistence across panels is ensured via `MetaObject`, which anchors recurring instances to a single canonical node throughout the narrative. Background elements without narrative function are instead delegated to the `NarrativeLocation` layer described in Section 3.5.

Speech and sound are represented by three complementary classes. A `Balloon` is a framed textual element transcribing an utterance produced by a character. Following Rissen [23], we refine the single-class model of [16] into a `balloonType` attribute with five values: `Speech` (explicit dialogue), `Thought` (inner monologue), `Whisper`, `Exclamation` (loud speech), and `Broadcast` (speech mediated through a device). Balloons further carry graphical attributes (`borderStyle`, `closureState`, `shape`, and `backgroundColor`) used to represent visual and tonal variation. Multi-balloon constructions are handled via a `linkedTo` property, with only the source balloon carrying a `Tail`. Each `Tail` points toward the utterance source through a cardinal `direction` and the coordinates of its tip; a single balloon may bear several tails when multiple characters share an utterance.

Complementing balloons, an `Onomatopoeia`, following Louis [19], is a textual element distinct from balloons, captions, and ordinary text lines that linguistically imitates a sound associated with the object or action it describes. The class carries a `content` datatype property storing the literal transcription and a `category` property covering ten canonical types. In contrast with previous work that treats onomatopoeia as isolated textual elements, we connect each instance to its sound source through a `hasOrigin` property pointing to the `Character` or `Object` responsible for the sound (*e.g.* a "VROOOM" onomatopoeia linked to a depicted motorcycle). Onomatopoeia may span several panels, in which case the same instance is linked to each of the `TimeLevels` it traverses.

A `Caption` is a textual container that conveys narrative voice rather than character speech or sound: it provides authorial narration, atmospheric cues, or inner monologue. Unlike a `Balloon`, which is anchored to a speaker or source via a `Tail`, a caption stands independent of any character and is typically rendered as a rectangular box. Captions are typed through a `captionType` property: `Narrative` for the voice of an external narrator that introduces or comments on an action (*e.g.* "And off our friends go for a new adventure..."); `Effect` for atmospheric, temporal, or situational indications (*e.g.* "The next morning..."); and `Internal` for a character's inner monologue rendered as a caption (*e.g.* "Those

words stayed engraved in my heart...”). Like any textual container, captions may carry `TextLine` and `TextSpan` elements.

Text lines and text spans. The `TextLine` class represents a single line of text as it physically appears on the plate. Using a dedicated class rather than a string datatype allows text lines to be annotated independently of their container: a text line may live inside a balloon, inside a caption, or freely within the panel itself, as is the case for text rendered on the façade of a building or on objects. The relationship is established through a `hasTextLine` property on the containing element. Within a `TextLine`, the `TextSpan` class provides a finer level of granularity, representing a word or group of words sharing a consistent typographic style. This distinction is justified by the fact that stylistic variations (*e.g.* bold, italic, or uppercase rendering) within a single line are often semantically loaded.

3.3 Intra-Panel Semantics

Even before committing to a full narrative structure, the visual layer described above supports a first level of semantics attached directly to individual panels and `TimeLevels`. The `Character` class carries an `emotion` attribute capturing the emotional state of a character in a specific local occurrence, drawn from the set of 28 common emotions distributed across the arousal-valence space of Russell's circumplex model [24].

A second semantic hook is provided by the `hasOrigin` property, which links each `Onomatopoeia` to its sound source in the scene, transforming what would otherwise be an isolated textual element into the graphical reflex of an event anchored to a participant, a grounding that proves essential for downstream tasks such as accessibility or summarisation.

The textual content of a balloon or caption may also refer to narrative entities absent from the current panel: an off-screen character, a verbally evoked object. The `mentions` object property addresses this by linking a `Balloon` or `Caption` directly to the relevant `MetaCharacter` or `MetaObject` instances. Anchoring mentions at the meta-level, rather than to a local occurrence, ensures they remain well-formed even when the referenced entity is not visually present.

Finally, actions and relations between characters and objects are represented through a dedicated `Interaction` class rather than a direct property between two entities. This design choice reflects the intrinsic complexity of interactions, which may involve several participants, an instrumental object, and a typed action. Drawing on standard semantic-role labelling [12], three roles are distinguished: `hasAgent` (who acts), `hasPatient` (who undergoes), and `hasInstrument` (with what). The action type is carried by `interactionType` as an infinitive verb. While `hasAgent` and `hasPatient` may point to characters or objects, `hasInstrument` and `hasLocative` are restricted to `Object`. Every `Interaction` is anchored to a single `TimeLevel`, ensuring that only co-present entities participate in it.

3.4 Narrative Structure

The visual and structural layer described in Section 3.1, together with the semantic layer of Section 3.3, account for the content of individual panels and levels, but do not capture how this content aggregates into a coherent story. A knowledge graph that can only enumerate characters, interactions, and balloons is, from a narrative standpoint, no different from a bag of disconnected images. Comics, like all narrative media, operate on two distinct temporal planes: the *fabula* (chronological order of events in the story world), and the *syuzhet* (order in which those events are presented to the reader). This distinction is particularly salient in comics, where flashbacks, flash-forwards, and parallel narration are common yet manifest as simple visual sequences unless explicitly encoded.

Existing frameworks. Two complementary theoretical sources inform the narrative layer. Cohn's Visual Narrative Grammar (VNG) [6, 9] proposes that comic panels are organised into hierarchical constituents governed by five functional roles: *Establisher* (E), which sets the referential context without advancing the action; *Initial* (I), which introduces a preparatory tension; *Peak* (P), the climactic moment heading the arc; *Release* (R), which resolves tension following a peak; and *Refiner* (r), which elaborates a neighbouring element without autonomous narrative force. VNG is structurally recursive, producing a constituency tree of arbitrary depth. Cohn cautions that it applies most strongly to wordless or image-dominant sequences: in multimodal narratives, VNG roles should be avoided when deleting the text would cause the imagery to lose its overall gist [9]. The second source is Chen's hierarchical knowledge graph [5], which stratifies narrative content into macro-events, events, and event segments.

Integration in the graph. We introduce `NarrativeConstituent` as the abstract superclass unifying all narrative units, organised into a five-level tree: `Story` → `Arc` → `Event` → `Scene` → `Segment`. This finer granularity, compared to Chen's three levels, separates editorial structure from cognitive event perception, provides a principled anchoring point for spatial information, and isolates the unique level at which VNG roles apply.

A `Story` represents the abstract *fabula* conveyed by a `Book`. It is subdivided into `Arc` instances defined by dramatic coherence rather than physical divisions. Each `Arc` contains `Event` instances unifying scenes around a common goal or causal thread. The `Scene` is defined by simultaneous stability of spatial, character, and action continuity, and is the level at which `hasNarrativeLocation` is anchored, avoiding redundant location assertions on lower constituents (see Section 3.5). At the leaf level, a `Segment` groups one or more `TimeLevel` instances via `instantiates` and is the only level bearing a `narrativeRole` attribute, constrained to {`Establisher`, `Initial`, `Peak`, `Release`, `Refiner`}.

The five subclasses are mutually disjoint, enforcing that each instance belongs to exactly one stratum. All levels are connected by a single `hasNarrativeUnit` object property linking any `NarrativeConstituent` to its immediate subconstituents. Every constituent may additionally carry a `narrativeEvent`

datatype property providing a free-text description, combining formal structure with human-readable natural language.

The critical bridge between the narrative and visual layers is the `instantiates` property, linking a `Segment` to the `TimeLevel` instances it subsumes. Through this single relation, every visual element attached to a `TimeLevel` is indirectly grounded in the narrative tree without redundant replication. This architecture ensures that querying the graph at the narrative level automatically reaches the visual layer, and that any visual element can be traced back to its narrative context, a foundational requirement for applications such as summarisation or accessibility.

3.5 Narrative Locations

To represent the places in which a narrative unfolds, we follow GOLEM in distinguishing two complementary levels [21]: `G12_Setting` (the global spatial, cultural, and social context of a work) and `G13_Narrative_Location` (the precise intra-diegetic environments in which events occur). We introduce a `NarrativeLocation` class that captures the narrative identity of a place (*"the forest"*, *"Paris"*) independently of how it is drawn in any particular panel. Locations are organised into a recursive hierarchy via the properties `hasSubLocation`: a room is contained in a building, a building in a street, a street in a city. This recursivity mirrors the recursivity of `NarrativeConstituents`: the two hierarchies are independent but intersect at the leaves, where each leaf constituent points to its narrative location via `hasNarrativeLocation`.

Next to this identity-level concept, we introduce a `SceneSetting` class that captures the *visual* appearance of a decor as rendered in a specific `TimeLevel`. Whereas `NarrativeLocation` states *"we are in the forest"*, `SceneSetting` states *"in this level we see a dense forest with a bird's nest in the tree, ..."*, and describes the background in natural language. Every `TimeLevel` carries its own `SceneSetting` via `hasSceneSetting`, and the scene setting is linked to the corresponding `NarrativeLocation` via `realizesLocation`. Several levels may share the same `SceneSetting`, which is typically the case for the panels of a polyptych.

3.6 Temporality

To encode both *syuzhet* and *fabula* we draw on the TimeML annotation scheme [22], which draws on Allen's interval relations [2] for natural language temporal annotation and organises temporal information around four primitives: EVENT, TIMEX3, SIGNAL, and LINK.

Reading order. The *syuzhet* is already encoded by the visual layer through `hasNextPlate` and `hasNextPanel`. No additional machinery is therefore required.

Story time. We introduce a `TemporalLink` class, the reified counterpart of TimeML's TLINK (one of three subtypes of LINK alongside SLINK and ALINK), relating two `NarrativeConstituents` through `tlinkType`. Its admissible values

are a minimal subset of TimeML's `relType`, restricted to the three cases most relevant to comic book narration: `BEFORE` (the source constituent precedes the target in the story world, covering flashbacks and flash-forwards), `SIMULTANEOUS` (source and target co-occur, covering parallel narration), and `IDENTITY` (source and target refer to the same story-world event seen from two different narrative positions). Annotation is only required at temporal ruptures; two consecutive constituents sharing the same *fabula* and *syuzhet* order require no link. This selective policy keeps the graph sparse while making non-linear structures explicit.

Temporal expressions. We introduce a `Timex` class, the counterpart of TimeML's TIMEX3, attached to `Effect`-type captions via `hasTimex`. It carries three properties: `timexText` records the original expression (*'the next morning"*, *'two years later"*), `timexType` specifies whether it refers to a `DATE`, `TIME`, or `DURATION`, and `timexValue` provides an optional ISO 8601 normalisation. Two dimensions remain out of scope and are discussed in Section 5: event durations when not textually stated, and visually implicit temporal markers (a shifting sky, a changing season, a clock in the background) conveyed by drawing alone.

4 Application on a comic book

To validate the ontology presented in Section 3 beyond its theoretical formulation, we instantiated it through the construction of a knowledge graph grounded in a manually annotated comic book, a process that took several days and which we intend to automate in future work. This experiment serves a dual purpose: it demonstrates that the ontological structure can accommodate the full complexity of a real sequential art work, and it produces a queryable knowledge base that can be evaluated against concrete information needs.

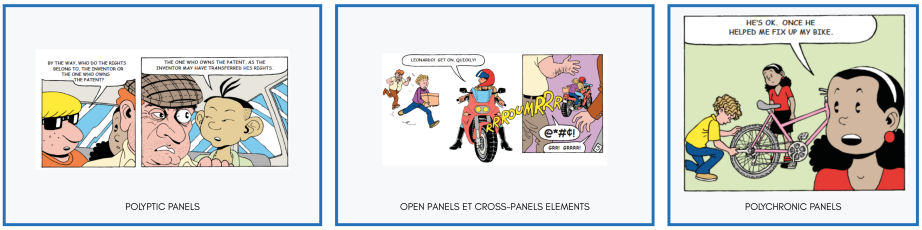

Fig. 2. Overview of the complex visual structure in *Patents*; full annotated graphs are provided in the supplementary repository.

4.1 Corpus selection

We selected Patents, a twelve-page bande dessinée produced by the World Intellectual Property Organization (WIPO) and freely available[5]. The work presents

[5] `https://www.wipo.int/publications/en/details.jsp?id=67` (accessed June 2026)

an interesting asymmetry for our purposes: while its institutional and pedagogical nature keeps its narrative arcs and character dynamics deliberately simple, it concentrates a technically representative range of visual conventions that pose the greatest challenge to existing annotation schemes. Specifically, it contains polychronic panels, in which several temporal strata coexist within a single frame; polyptych sequences, in which a continuous background is fragmented across multiple panels while characters appear at successive instants within it; multi-balloon constructions linking several speech bubbles into a single continuous utterance; onomatopoeia that physically span two adjacent panels; and flashback structures that create a non-trivial divergence between the *fabula* and the *syuzhet*. This combination makes Patents a well-suited first test case for the structural and visual dimensions of the ontology, exercising precisely the aspects of the framework that existing annotation schemes handle least well, while its deliberately simple narrative structure keeps the evaluation focused and interpretable at this stage of development. Figure 2 illustrates representative examples of the phenomena encoded by the annotation. (e.g. polyptic and polychronic panels, cross-panels elements).

4.2 Manual annotation and resulting graph

The annotation was carried out manually, encoding each plate, panel, level, character occurrence, and all items as a set of RDF triples conforming to 9$^{\text{th}}$ArtOnto. The resulting knowledge graph comprises **1,414** nodes and **2,307** arcs, representing the complete set of visual, semantic, and narrative elements identified across the twelve pages of the work. [6]

This structured representation enables the reconstruction of the narrative timeline from the annotated sequence of panels: by following the `TemporalLink` relations between narrative constituents, the chronological order of events in the story world (the *fabula*) can be recovered from a presentation order (the *syuzhet*) that includes non-linear structures such as flashbacks. Beyond temporal reconstruction, the graph supports a range of structured queries over the full content of the work, such as retrieving all speech balloons uttered by a given character, enumerating the panels in which two specific characters appear simultaneously, or identifying all onomatopoeia associated with a particular sound source. Such query patterns, which would be impossible to express over an unstructured document, confirm that the graph constitutes a genuine knowledge base over the work rather than a mere transcription of its visual content.

5 Limits

The framework presented here constitutes a first iteration of 9$^{\text{th}}$ArtOnto, and as such it deliberately restricts its scope to the dimensions of sequential art that

[6] The annotated graph, the ontology and a visualisation interface is available here : `https://gitlab.univ-lr.fr/dataset/comics-graph-viewer`

can be modelled with reasonable confidence at this stage of development. Several aspects of the medium remain outside the current perimeter and represent natural directions for subsequent versions.

A first category of limitations concerns the modelling of characters, both at the level of individual identity and of social relations. The current `MetaCharacter`/`Character` pattern tracks identity across panels by accommodating variations in posture, expression, scale, or drawing style as local manifestations of a persistent entity; yet it cannot handle cases where a character's appearance changes in narratively meaningful ways, such as a character depicted at different ages or wearing a disguise, whose identity continuity remains invisible to the visual surface. At the relational level, 9th ArtOnto captures interactions as local, panel-level events through the `Interaction` class, but does not represent the persistent social structures that bind characters across an entire work: friendship, rivalry, familial ties, hierarchical authority, and the evolution of these bonds through narrative events. These dimensions, which are well handled in narrative ontologies such as GOLEM through its *Social Relationship* and *Relationship Role* modules, are deliberately out of scope here, where the primary ambition is to ground the knowledge graph in visually annotated content rather than to infer higher-level social dynamics.

A second gap concerns temporality: while `TemporalLink` and `Timex` handle non-linear narrative structures and explicit temporal expressions, two dimensions remain unrepresented. The first is the duration of narrative events when not made explicit by an `Effect-type` caption, leaving the model unable to express how long a `NarrativeConstituent` lasts when only the artwork conveys this. The second is the encoding of visually implicit temporal markers (*e.g.* a sky transitioning from day to night, a changing season, a clock in the background), which would require integrating image analysis capabilities outside the present ontological scope.

A third limitation concerns the representativeness of the corpus used for validation. *Patents* is an institutional, pedagogical work whose narrative structure is deliberately simple and linear. The five-level constituency hierarchy and the `TemporalLink` machinery were therefore applied to a work with limited divergence between *fabula* and *syuzhet*, leaving the boundaries between narrative levels largely untested. Extending the framework to longer, author-driven works spanning different genres and visual styles remains a necessary step toward assessing whether these distinctions are stable and expressive enough to capture the full range of narrative structures found in sequential art.

Finally, one aspect not yet addressed is the External Compositional Structure of pages, as theorised by Cohn [7]: the way panels are spatially organised on a plate independently of their internal content, through rows, columns, grids, overlapping insets, or more complex hierarchical arrangements. Distinct from both panel content and the reading order encoded by `hasNextPanel`, this page-level layout is narratively and rhetorically significant, as the spatial organisation of a plate is itself an expressive act shaping the reader's experience of rhythm and pace.

6 Conclusion and future works

We have introduced *9th ArtOnto*, an OWL 2 ontology providing a first formal framework for the joint representation of the visual, structural, and narrative dimensions of sequential art works. The ontology captures every graphical constituent of a work from the plate down to the individual text span, encodes character emotions, interactions, mentions, and onomatopoeia grounding, and structures panels into a five-level constituency hierarchy that enables the reconstruction of both the *fabula* and the *syuzhet*, including non-linear devices such as flashbacks. The framework was preliminarily illustrated through its application to *Patents*, offering an initial indication that the ontology can accommodate the complexity of a real sequential art work within a single, queryable knowledge graph, though a more thorough validation on a broader corpus remains necessary to fully assess its expressiveness and stability.

Beyond the ontology itself, the resulting knowledge graph opens the way to a range of applications. Automatic summarisation could exploit the narrative constituency hierarchy to generate plot synopses at varying levels of granularity. Context-aware translation could leverage character identity, speech attribution, and scene-level information to resolve ambiguities that purely textual approaches miss. Natural language retrieval would allow users to query comic book corpora through structured questions, for instance, identifying all scenes in which two specific characters interact under a given narrative condition. More broadly, the structured representation produced by the framework could support comparative analysis across works, genres, or authors, as well as accessibility tools aimed at generating descriptive captions for visually impaired readers.

The automatic population of knowledge graphs conforming to *9th ArtOnto* from raw comic book pages represents one of the most promising research directions openend by this work. Tasks such as character identity resolution, narrative segmentation, and temporal modelling require a level of understanding that current systems do not yet reliably achieve, and bridging this gap constitutes the primary research direction we intend to pursue.

Acknowledgments

This work was supported by the French National Research Agency (ANR) under the project EnACA (grant number ANR-25-CE38-6411).

References

1. Aalberg, T., Riva, P., Žumer, M.: Lrmoo: object-oriented definition and mapping from the ifla library reference model (2024)
2. Allen, J.F., Hayes, P.J.: A Common-Sense Theory of Time (1983)
3. Augereau, O., Iwata, M., Kise, K.: A Survey of Comics Research in Computer Science. Journal of Imaging **4**(7), 87 (Jun 2018)

4. Busi Rizzi, G.: Always at home in the past: exploring nostalgia in the graphic novel. Ph.D. thesis, Università di Bologna. Dipartimento di Lingue, Letterature e Culture Moderne ... (2018)
5. Chen, Y.C.: Hierarchical Knowledge Graphs for Story Understanding in Visual Narratives (Nov 2025), arXiv:2506.10008 [cs]
6. Cohn, N.: The limits of time and transitions: challenges to theories of sequential image comprehension. Studies in Comics **1**(1), 127–147 (Apr 2010)
7. Cohn, N.: Navigating Comics: An Empirical and Theoretical Approach to Strategies of Reading Comic Page Layouts. Frontiers in Psychology **4** (2013)
8. Cohn, N.: Visual Narrative Structure. Cognitive Science **37**(3), 413–452 (2013)
9. Cohn, N.: Your Brain on Comics: A Cognitive Model of Visual Narrative Comprehension. Topics in Cognitive Science **12**(1), 352–386 (Jan 2020)
10. Damiano, R., Lombardo, V., Pizzo, A.: The ontology of drama. Applied Ontology **14**(1), 79–118 (Feb 2019)
11. Doerr, M.: The cidoc conceptual reference module: an ontological approach to semantic interoperability of metadata. AI magazine **24**(3), 75–75 (2003)
12. Fillmore, C.J.: The case for case (1967)
13. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with dolce. In: International conference on knowledge engineering and knowledge management. pp. 166–181. Springer (2002)
14. Grau, B.C., Horrocks, I., Motik, B., Parsia, B., et al.: OWL 2: The next step for OWL. Journal of Web Semantics **6**(4), 309–322 (Nov 2008)
15. Guérin, C.: Proposition d'un cadre pour l'analyse automatique, l'interprétation et la recherche interactive d'images de bande dessinée. Ph.D. thesis (2014)
16. Guérin, C., Rigaud, C., Bertet, K., Revel, A.: An ontology-based framework for the automated analysis and interpretation of comic books' images. Information Sciences **378**, 109–130 (Feb 2017)
17. Hogan, A., et al.: Knowledge Graphs. ACM Computing Surveys **54**(4)
18. Ji, S., et al.: A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. IEEE Transactions on Neural Networks and Learning Systems **33**(2)
19. Louis, J.B.: Extraction d'éléments textuels complexes: Application à la détection et à la reconnaissance des onomatopées dans les bandes dessinées. Ph.D. thesis (2025)
20. Meghini, C., Bartalesi, V., Metilli, D.: Representing narratives in digital libraries: The narrative ontology. Semantic Web **12**(2), 241–264 (Jan 2021)
21. Pianzola, F., Cheng, L., Pannach, F., Yang, X., Scotti, L.: The GOLEM Ontology for Narrative and Fiction. Humanities **14**(10), 193 (Oct 2025)
22. Pustejovsky, J., et al.: Timeml: Robust specification of event and temporal expressions in text. New directions in question answering **3**, 28–34 (2003)
23. Rissen, P.: A Comics Ontology (Apr 2012)
24. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology **39**(6), 1161–1178 (Dec 1980)
25. Scott McCloud: Understanding Comics: The Invisible Art (1993)
26. Sharma, R., Kukreja, V.: CPD: Faster RCNN-based DragonBall Comic Panel Detection. In: 12th International Conference on Communication Systems and Network Technologies (CSNT). pp. 786–790. IEEE, Bhopal, India (Apr 2023)
27. Vivoli, E., Campaioli, I., et al.: Comics Datasets Framework: Mix of Comics datasets for detection benchmarking (Jul 2024), arXiv:2407.03540 [cs]
28. Walsh, J.A.: Comic Book Markup Language: An Introduction and Rationale. Digital Humanities Quarterly **6**(1) (May 2012)