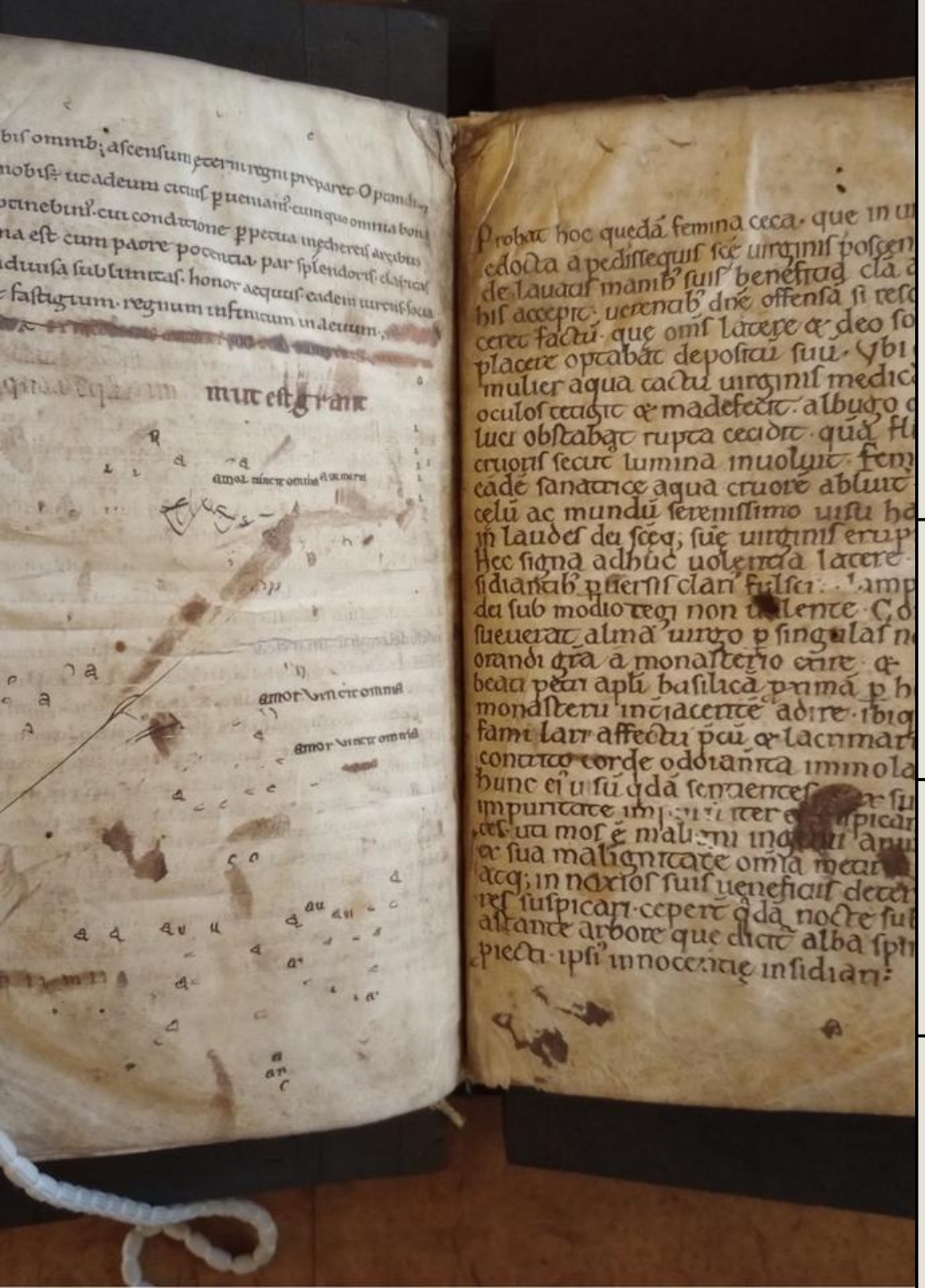


MANUSCRIPT-AI ONTOLOGY

A structured ontology for medieval manuscript description, interoperability, and retrieval

SHARI BOODTS (RADBOUD UNIVERSITY)
shari.boodts@ru.nl

GIULIA BIAGIONI (TNO)
giulia.biagioni@tno.nl

ERC Proof of Concept Lump Sum Grant 2023 (ERC-2023-PoC)

Why this ontology?

Manuscript data is often heterogeneous, fragmented, and difficult to integrate

Cultural heritage institutions need consistent models for cataloguing and reuse

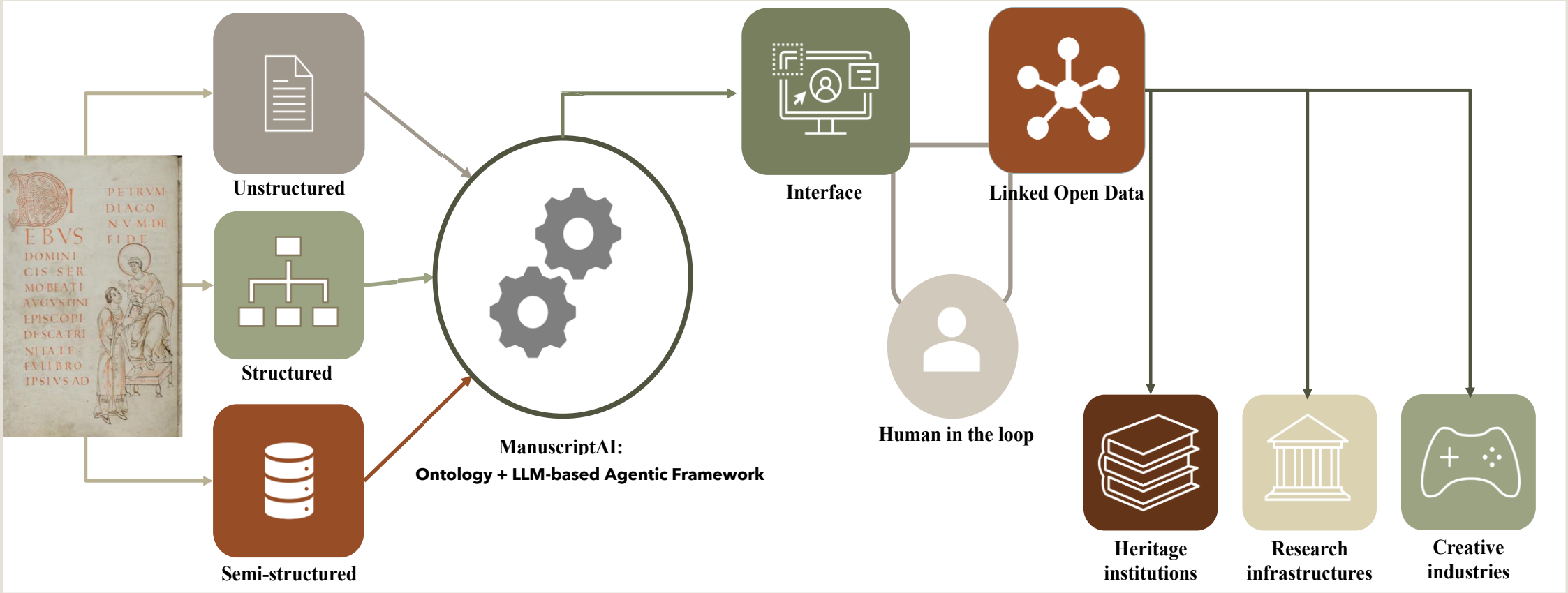
Researchers need richer, more precise, machine-readable manuscript descriptions

Manuscript-AI Ontology addresses these needs through a structured semantic framework

Why this ontology?

Key goal: improve cataloguing, retrieval, preservation, and interoperability of manuscript data

MANUSCRIPT-AI – Workflow



METHODOLOGY

Structural rationale: ontology design for neuro-symbolic knowledge graph construction

The model was designed for a workflow in which agents in the ManuscriptAI framework identify relevant semantic categories in text and map them to concepts in the ontology. A rule-based layer then reconstructs the RDF knowledge graph from this controlled mapping.

We therefore adopted a model construction designed to reduce the complexity of the mapping process and to provide a stable schema for the rule-based construction of the final knowledge graph.

Conceptual rationale: grounded in the specificity of medieval manuscripts

The ontology was conceptually informed by the specific descriptive requirements of medieval manuscripts.

This grounding draws on **Muzerelle's** ***Vocabulaire codicologique*** and on **domain expertise**.

It guided the selection and organization of the main classes, properties, and manuscript features represented in the model.

What the ontology covers

Coverage

- Events related to the creation of manuscript descriptions
- Manuscripts as cultural-heritage objects
- The internal organization of manuscripts through loci
- Entities, places, works, and physical features associated with manuscripts
- Detailed textual, paratextual, layout, material, and codicological features at locus level

Core Classes

1- Creation Manuscript Description

2- Manuscript

3- Locus

4- Locus Feature

What the ontology covers

Coverage

- Events related to the creation of manuscript descriptions
- Manuscripts as cultural-heritage objects
- The internal organization of manuscripts through loci
- Entities, places, works, and physical features associated with manuscripts
- Detailed textual, paratextual, layout, material, and codicological features at locus level

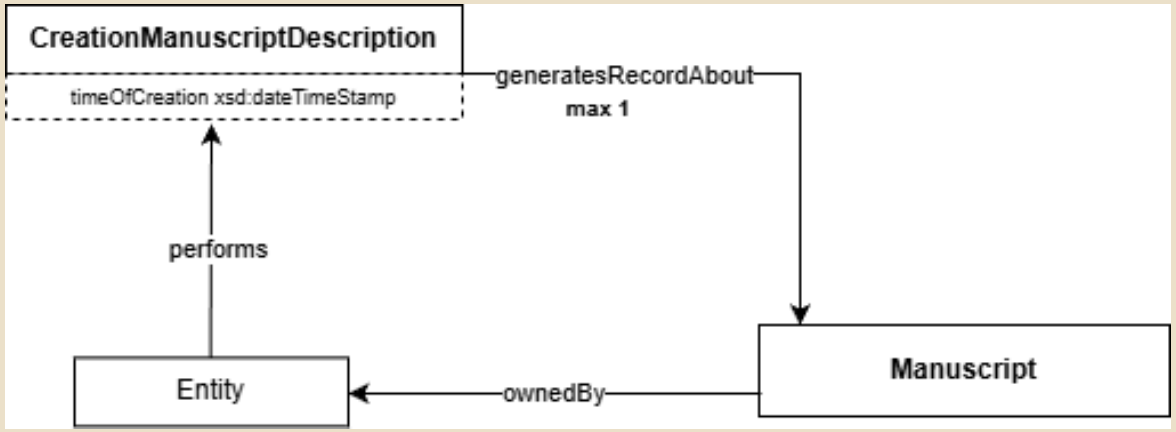
Core Classes

1- Creation Manuscript Description

2- Manuscript

3- Locus

4- Locus Feature



Definition:

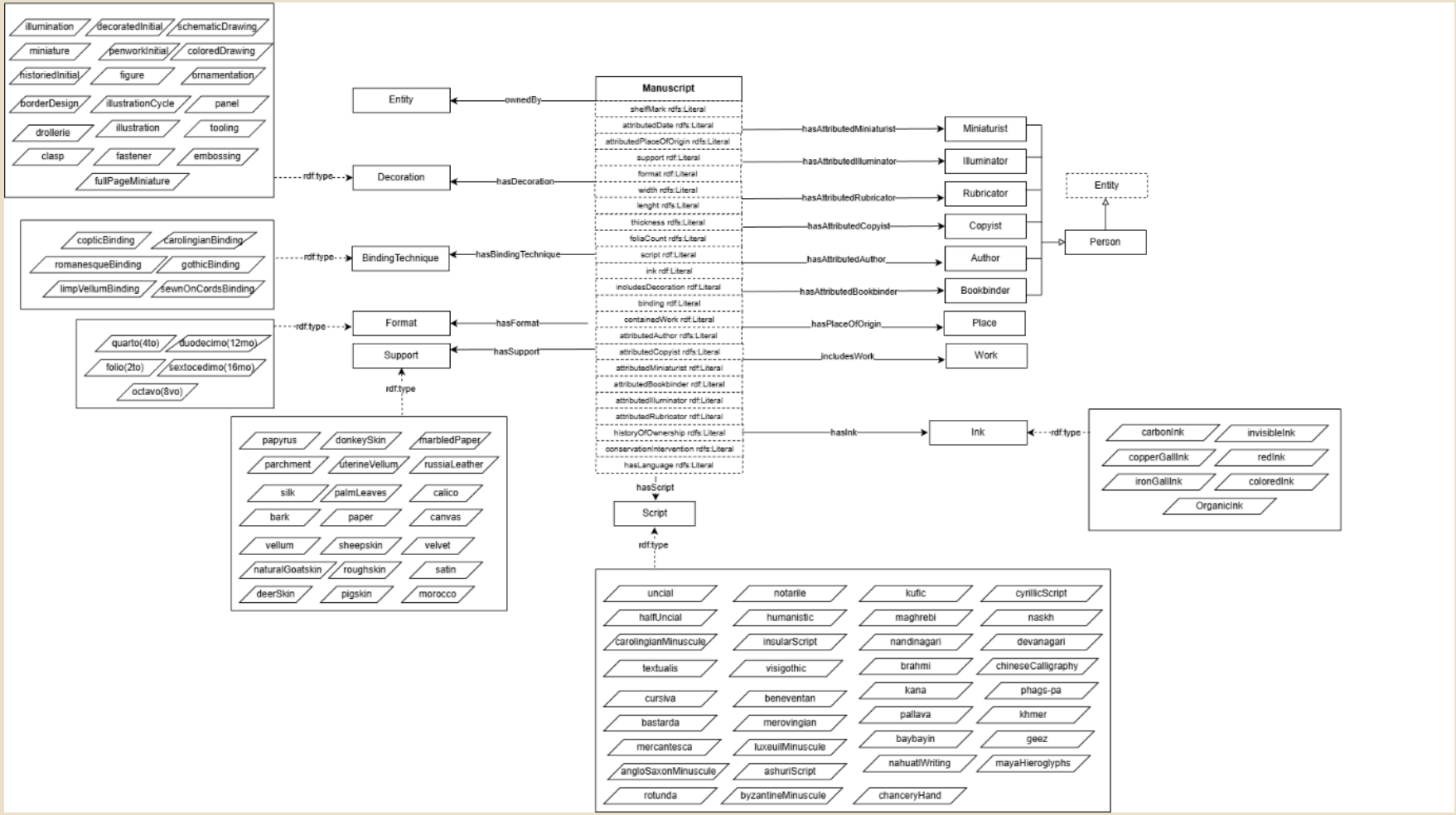
Represents the event in which a manuscript description is created by an entity

The entity may be a person, institution, or agent

The event happens at a specific time

Each event generates a record about one manuscript

Importance: Captures who described a manuscript, when, and in relation to which object

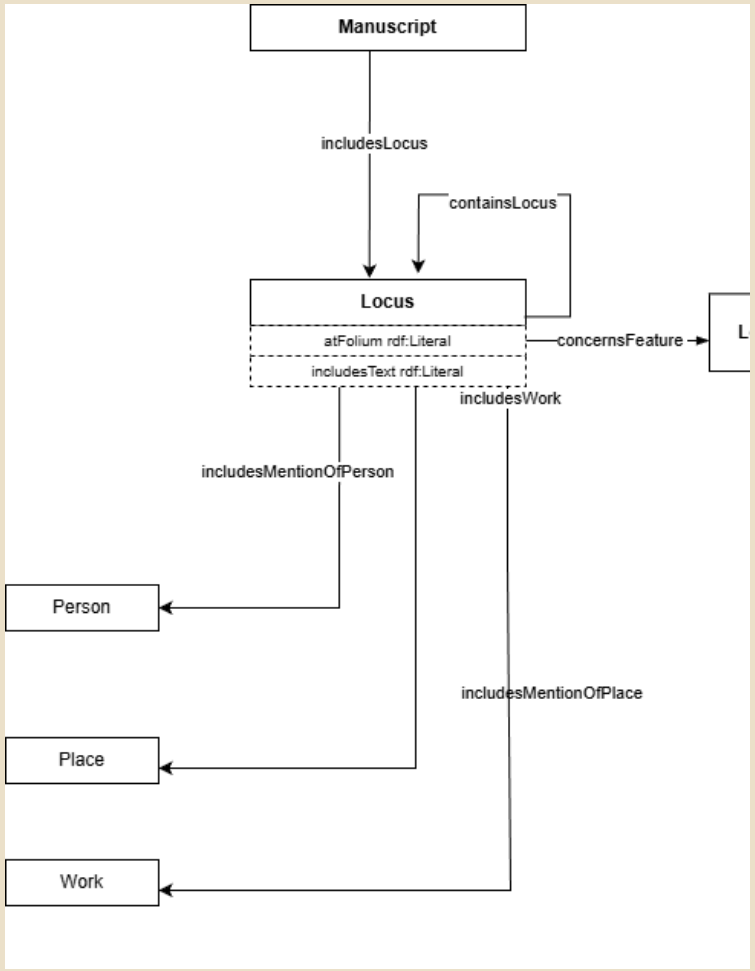


How Manuscript is described to support both simple cataloguing and richer linked-data practices:

Datatype properties for literal description

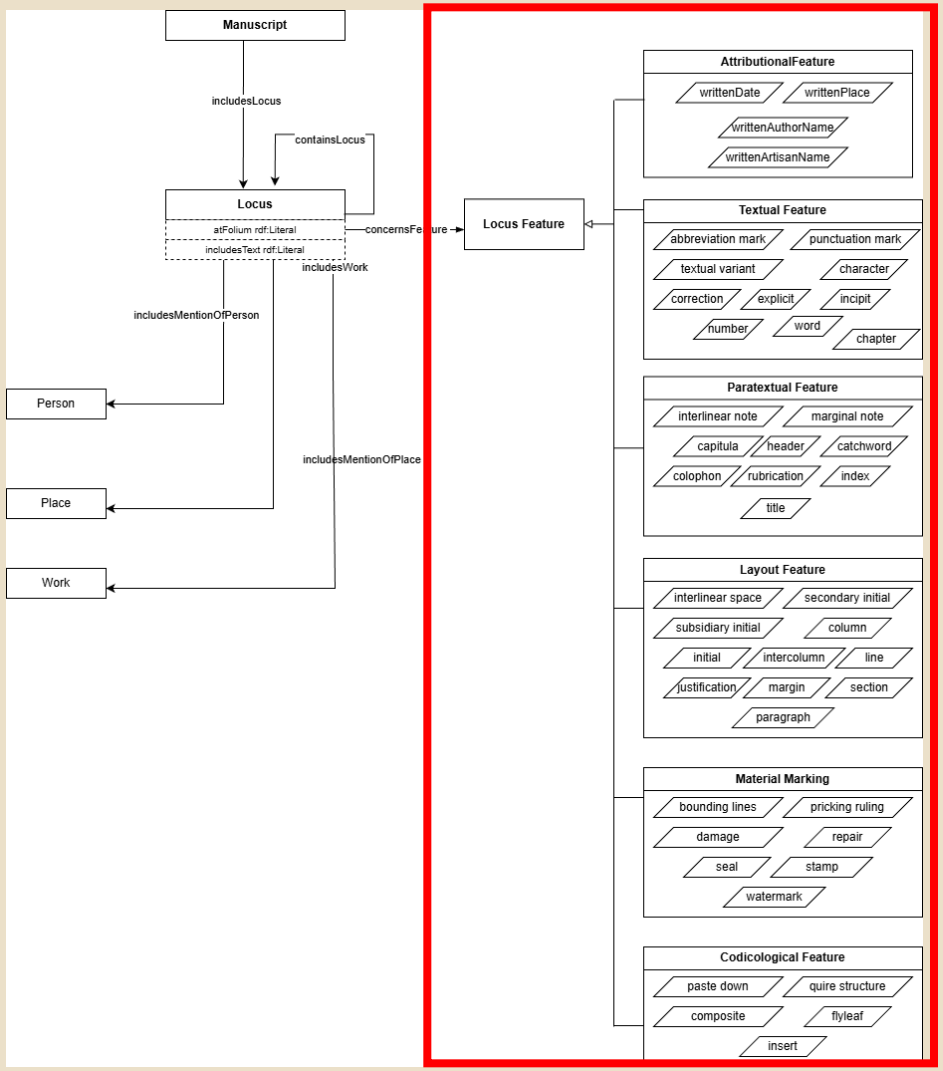
Object properties for semantic linking

Controlled instances for recurring manuscript features



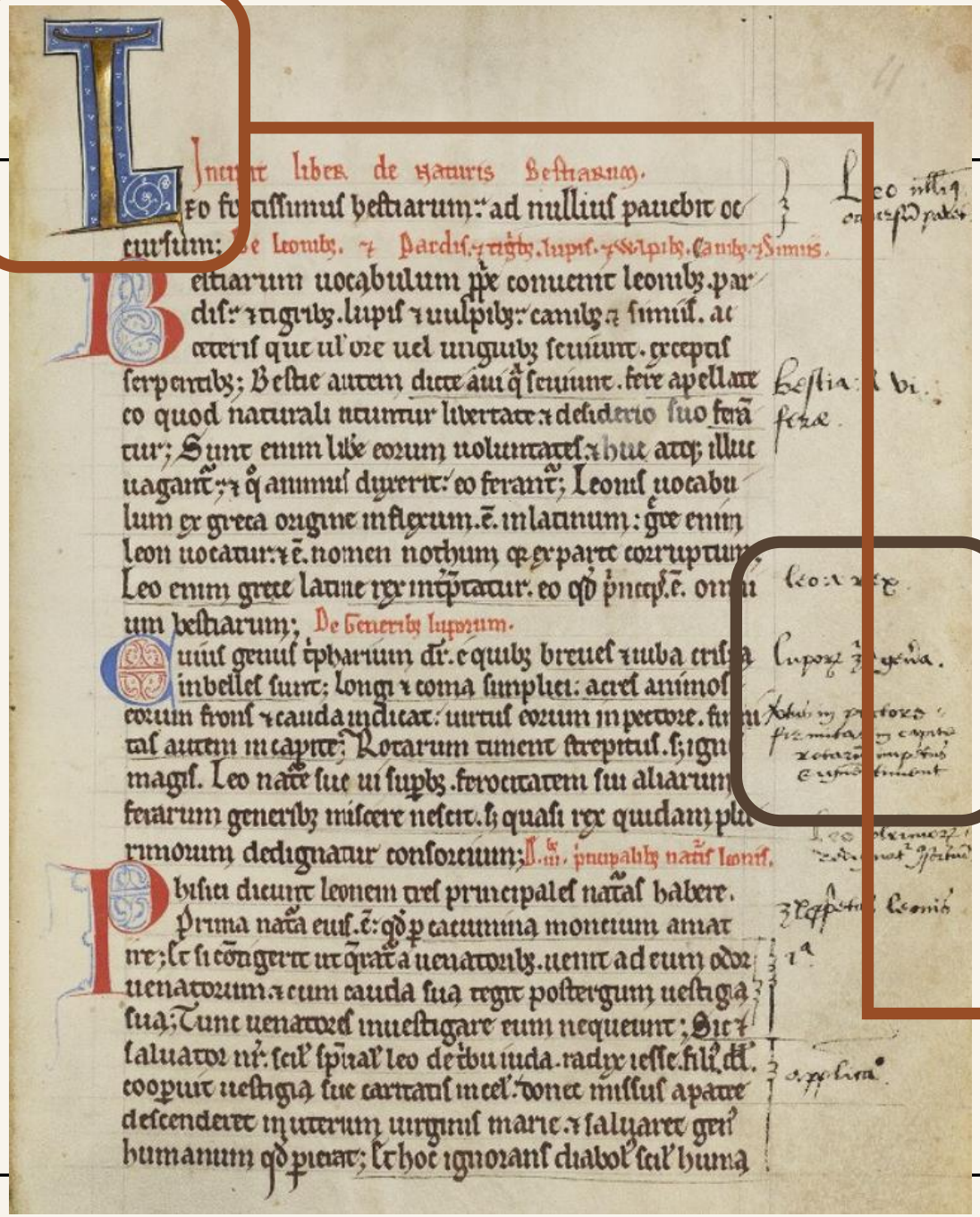
Definition: A Locus is an identifiable internal portion of a manuscript

Role: Enables description below the manuscript level



LocusFeature captures phenomena observable in a specific locus

Main subclasses:
AttributionalFeature
TextualFeature
ParatextualFeature
LayoutFeature
MaterialMarking
CodicologicalFeature



Locus Feature

Textual Feature

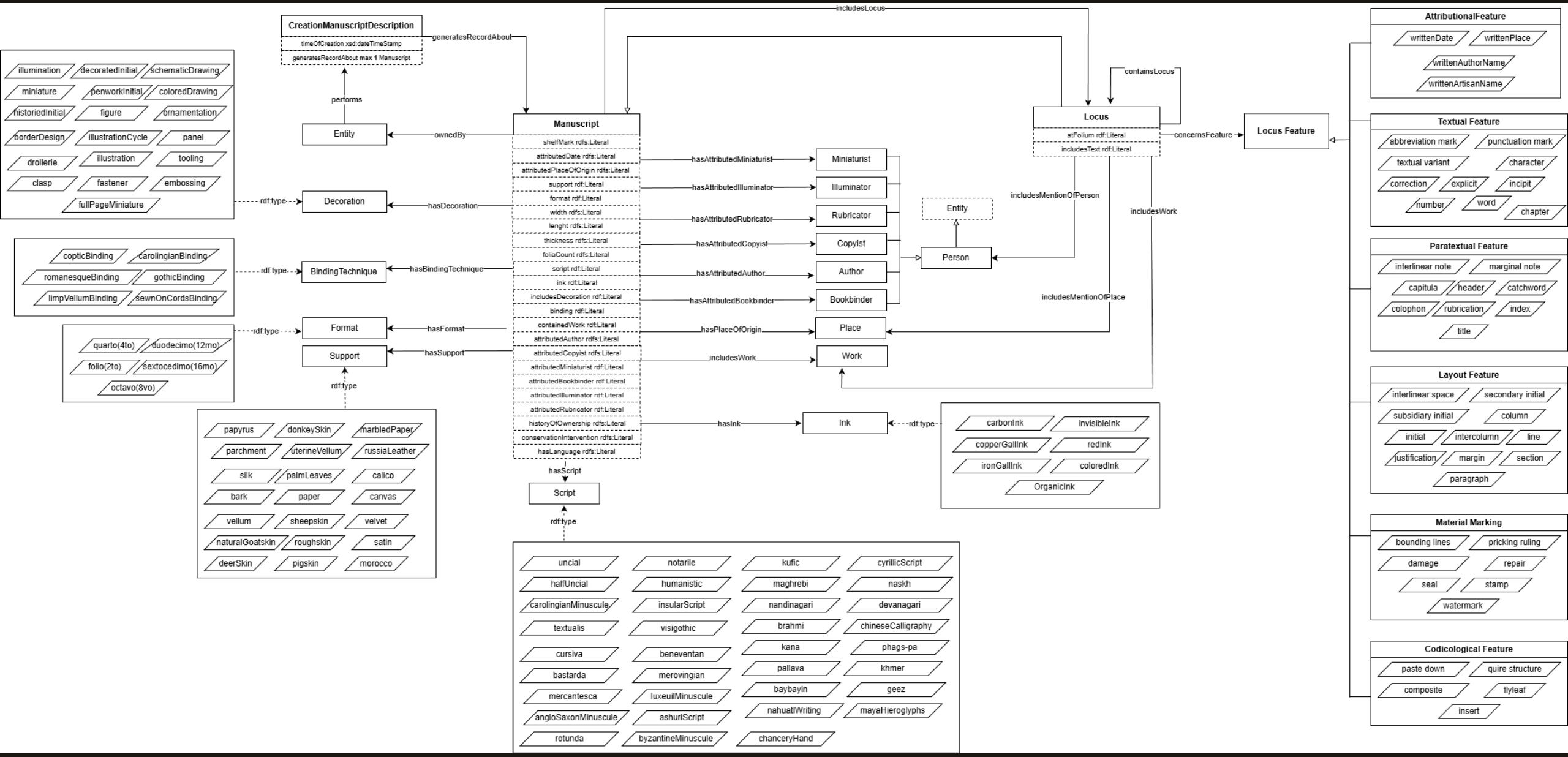
- abbreviation mark
- punctuation mark
- textual variant
- character
- correction
- explicit
- incipit
- number
- word

Paratextual Feature

- marginal note
- capitula
- header
- catchword
- colophon
- rubrication
- index
- title

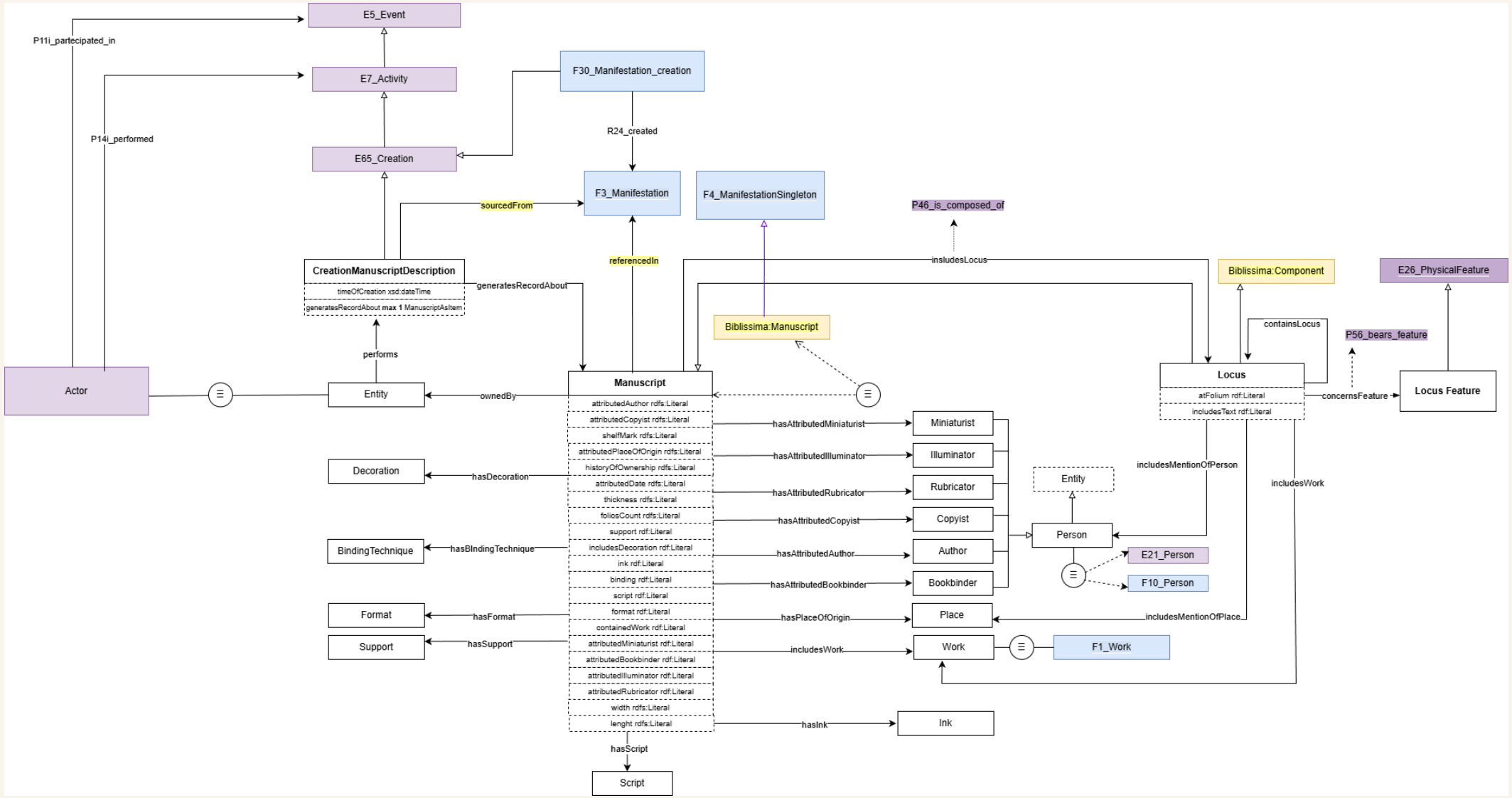
Layout Feature

- interlinear space
- secondary initial
- subsidiary initial
- column
- initial
- intercolumn
- line
- justification
- margin
- section
- paragraph


CreationManuscriptDescription
timeOfCreation xsd:dateTimeStamp
generatesRecordAbout max 1 Manuscript
generatesRecordAbout
includesLocus
performs
Entity
ownedBy
Manuscript
shelfMark rdfs:Literal
attributedDate rdfs:Literal
attributedPlaceOfOrigin rdfs:Literal
support rdf:Literal
format rdf:Literal
width rdfs:Literal
lenght rdfs:Literal
thickness rdfs:Literal
foliaCount rdfs:Literal
script rdf:Literal
ink rdf:Literal
includesDecoration rdf:Literal
binding rdf:Literal
containedWork rdf:Literal
attributedAuthor rdfs:Literal
attributedCopyist rdfs:Literal
attributedMiniaturist rdf:Literal
attributedBookbinder rdf:Literal
attributedIlluminator rdf:Literal
attributedRubricator rdf:Literal
historyOfOwnership rdfs:Literal
conservationIntervention rdfs:Literal
hasLanguage rdfs:Literal
hasAttributedMiniaturist
hasAttributedIlluminator
hasAttributedRubricator
hasAttributedCopyist
hasAttributedAuthor
hasAttributedBookbinder
hasPlaceOfOrigin
includesWork
hasInk
hasScript
Script
rdf:type
Miniaturist
Illuminator
Rubricator
Copyist
Author
Bookbinder
Entity
Person
Place
Work
Ink
rdf:type
carbonInk
invisibleInk
copperGallInk
redInk
ironGallInk
coloredInk
OrganicInk
Locus
atFolium rdf:Literal
includesText rdf:Literal
containsLocus
concernsFeature
includesMentionOfPerson
includesWork
includesMentionOfPlace
Locus Feature
illumination
decoratedInitial
schematicDrawing
miniature
penworkInitial
coloredDrawing
historiedInitial
figure
ornamentation
borderDesign
illustrationCycle
panel
drollerie
illustration
tooling
clasp
fastener
embossing
fullPageMiniature
rdf:type
Decoration
hasDecoration
copticBinding
carolingianBinding
romanesqueBinding
gothicBinding
limpVellumBinding
sewnOnCordsBinding
rdf:type
BindingTechnique
hasBindingTechnique
quarto(4to)
duodecimo(12mo)
folio(2to)
sextocedimo(16mo)
octavo(8vo)
rdf:type
Format
hasFormat
Support
hasSupport
rdf:type
papyrus
donkeySkin
marbledPaper
parchment
uterineVellum
russiaLeather
silk
palmLeaves
calico
bark
paper
canvas
vellum
sheepskin
velvet
naturalGoatskin
roughskin
satin
deerSkin
pigskin
morocco
rdf:type
uncial
notarile
kufic
cyrillicScript
halfUncial
humanistic
maghrebi
naskh
carolingianMinuscule
insularScript
nandinagari
devanagari
textualis
visigothic
brahmi
chineseCalligraphy
cursiva
beneventan
kana
phags-pa
bastarda
merovingian
pallava
khmer
mercantesca
luxeuilMinuscule
baybayin
geez
angloSaxonMinuscule
ashuriScript
nahuatlWriting
mayaHieroglyphs
rotunda
byzantineMinuscule
chanceryHand
AttributionalFeature
writtenDate
writtenPlace
writtenAuthorName
writtenArtisanName
Textual Feature
abbreviation mark
punctuation mark
textual variant
character
correction
explicit
incipit
number
word
chapter
Paratextual Feature
interlinear note
marginal note
capitula
header
catchword
colophon
rubrication
index
title
Layout Feature
interlinear space
secondary initial
subsidiary initial
column
initial
intercolumn
line
justification
margin
section
paragraph
Material Marking
bounding lines
pricking ruling
damage
repair
seal
stamp
watermark
Codicological Feature
paste down
quire structure
composite
flyleaf
insert

The Manuscript-AI Ontology aims to align, where appropriate, with established semantic frameworks in the cultural-heritage and bibliographic domains, particularly **CIDOC-CRM**, **FRBRoo/LRMoo**, and **Biblissima**, in order to support interoperability, data integration, and semantic reuse

HOW MANUSCRIPT-AI ALIGNS WITH EXTERNAL MODELS



Future work will focus on:

A) Strengthening the conceptual and technical alignment of the Manuscript-AI Ontology with **CIDOC CRM** This includes refining the model's design in closer accordance with **CIDOC CRM** principles and structures

B) In the longer term, we aim to explore whether the ontology could be proposed as an **externally maintained model** recognized as compatible with the **CIDOC CRM** framework

THANK YOU!