MANUSCRIPT-AI

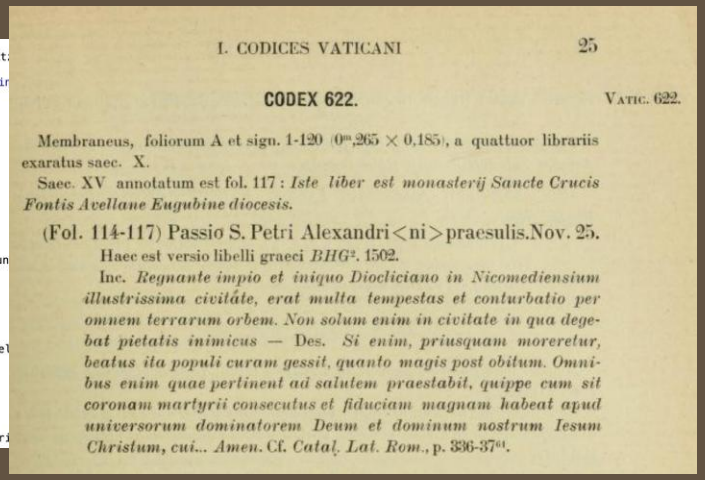
Pilot Results on Manuscript-AI Ontological Framework

Giulia Biagioni (TNO)

Shari Boodts (Radboud University)

- **EU Horizon Proof of concept** to evaluate the potential for automated integration, mapping and representation of diverse cultural heritage data collections into a machine-understandable format, using **medieval manuscripts** as a case study
- April 2024 to March 2026 (23 months)
- Spin-off of Horizon ERC Starting Grant 'Patristic Sermons in the Middle Ages' (PASSIM) 2019-2024.
- Collaboration between **Radboud University** and the **Dutch Organization for Applied Scientific Research** (TNO).

From Parchment to Pixels: The Digital Spectrum of Manuscripts



I. CODICES VATICANI 25

CODEX 622. VATIC. 622.

Membraneus, foliorum A et sign. 1-120 (0m,265 × 0,185), a quattuor librariis exaratus saec. X.

Saec. XV annotatum est fol. 117 : *Iste liber est monasterij Sancte Crucis Fontis Avellane Eugubine diocesis.*

(Fol. 114-117) Passio S. Petri Alexandri<ni>praesulis.Nov. 25.

Haec est versio libelli graeci *BHG*². 1502.

Inc. *Regnante impio et iniquo Diocliciano in Nicomediensium illustrissima civitate, erat multa tempestas et conturbatio per omnem terrarum orbem. Non solum enim in civitate in qua degebat pietatis inimicus — Des. Si enim, priusquam moreretur, beatus ita populi curam gessit, quanto magis post obitum. Omnibus enim quae pertinent ad salutem praestabit, quippe cum sit coronam martyrii consecutus et fiduciam magnam habeat apud universorum dominatorem Deum et dominum nostrum Iesum Christum, cui... Amen.* Cf. *Catal. Lat. Rom.*, p. 336-37[61].

European Manuscript collections are subjected to different levels of datafication:

- Some collections are fully digitized and datafied as linked (open) data.
- Other are organized in semi-structured relational databases and isolated into digital silos.
- Some collections are described and recorded only within paper registries.
- Certain collections remain almost completely undocumented or noted only in specialized research and scientific articles (implicitly recorded).

Issues

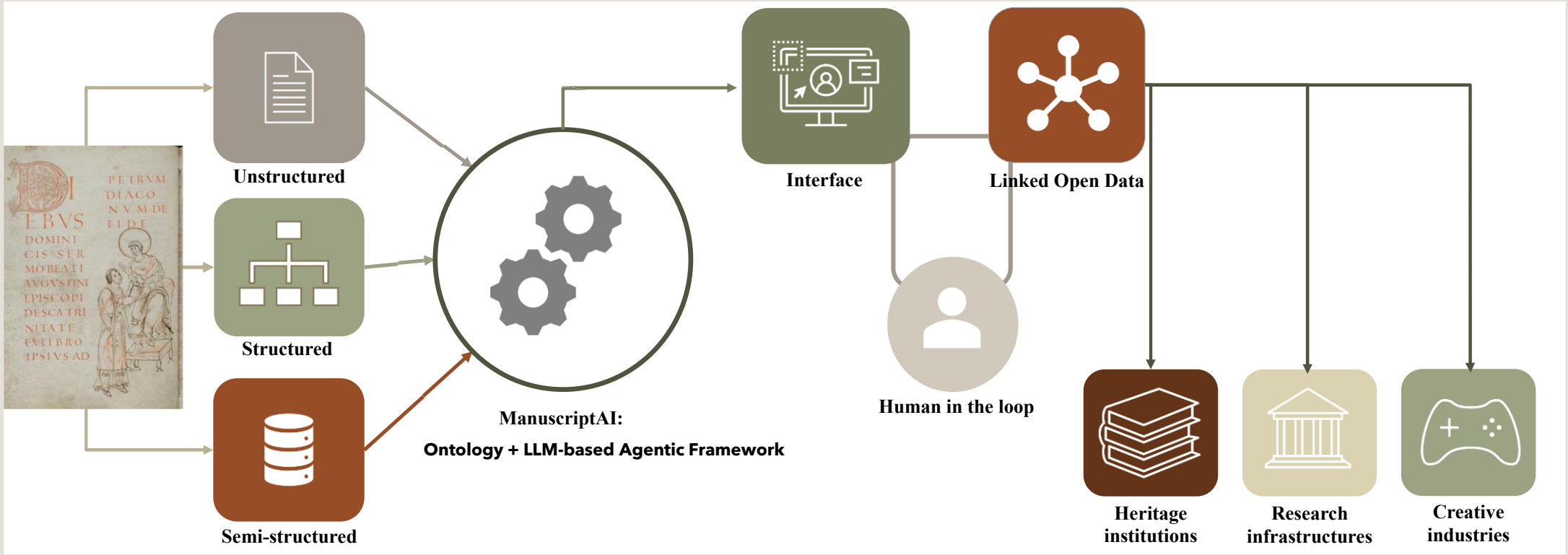
Impossibility to preserve human history both in both in memory and actuality

Cross-referencing analysis across various data collections, straightforward retrieval of information, is not feasible.

Impossibility to protect cultural heritage from neglect and theft, criminal use and conflict during war-time

Uneven datafication of collections; well-resourced collections gain visibility, while those less-funded risk obscurity.

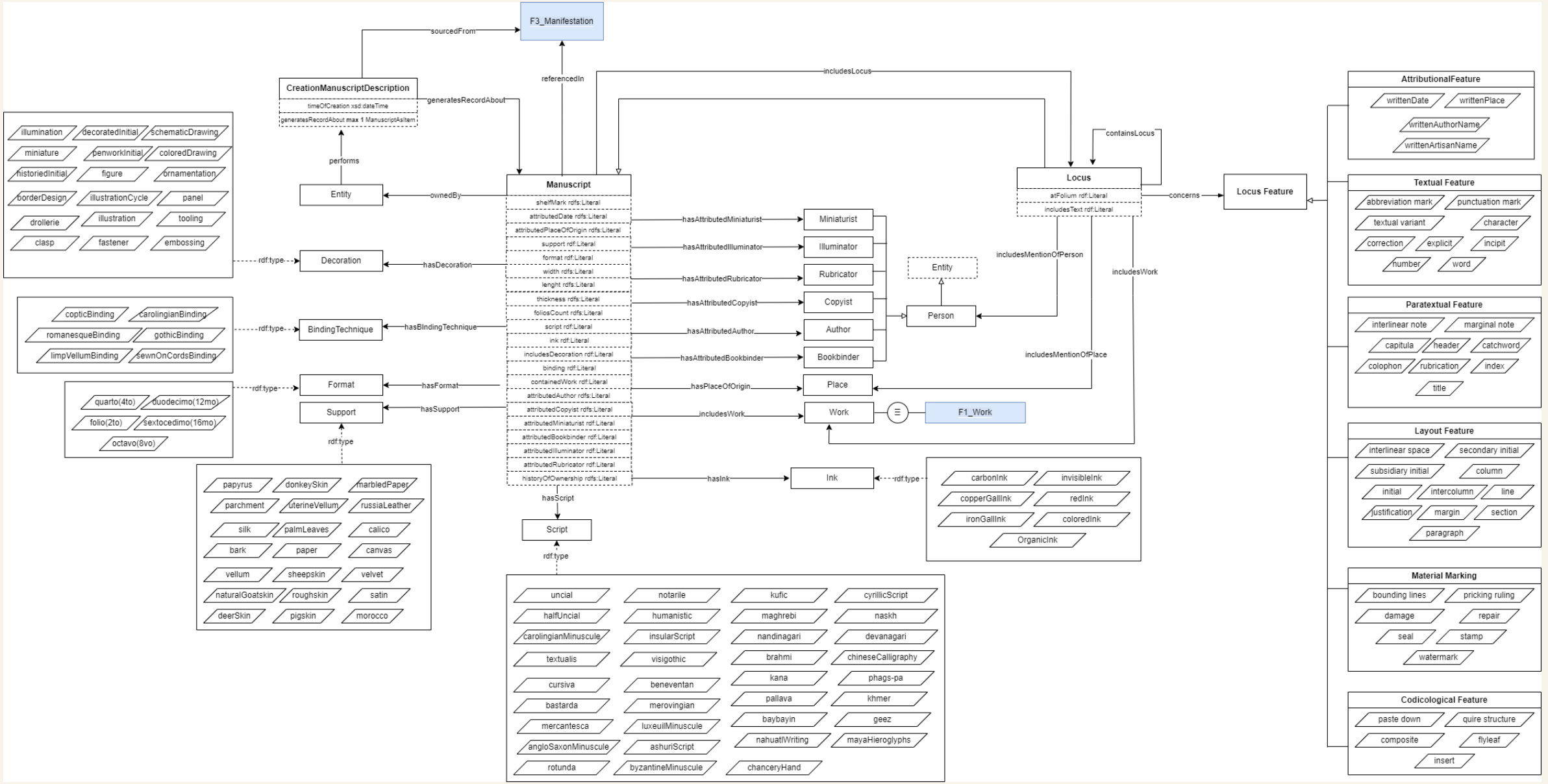
MANUSCRIPT-AI

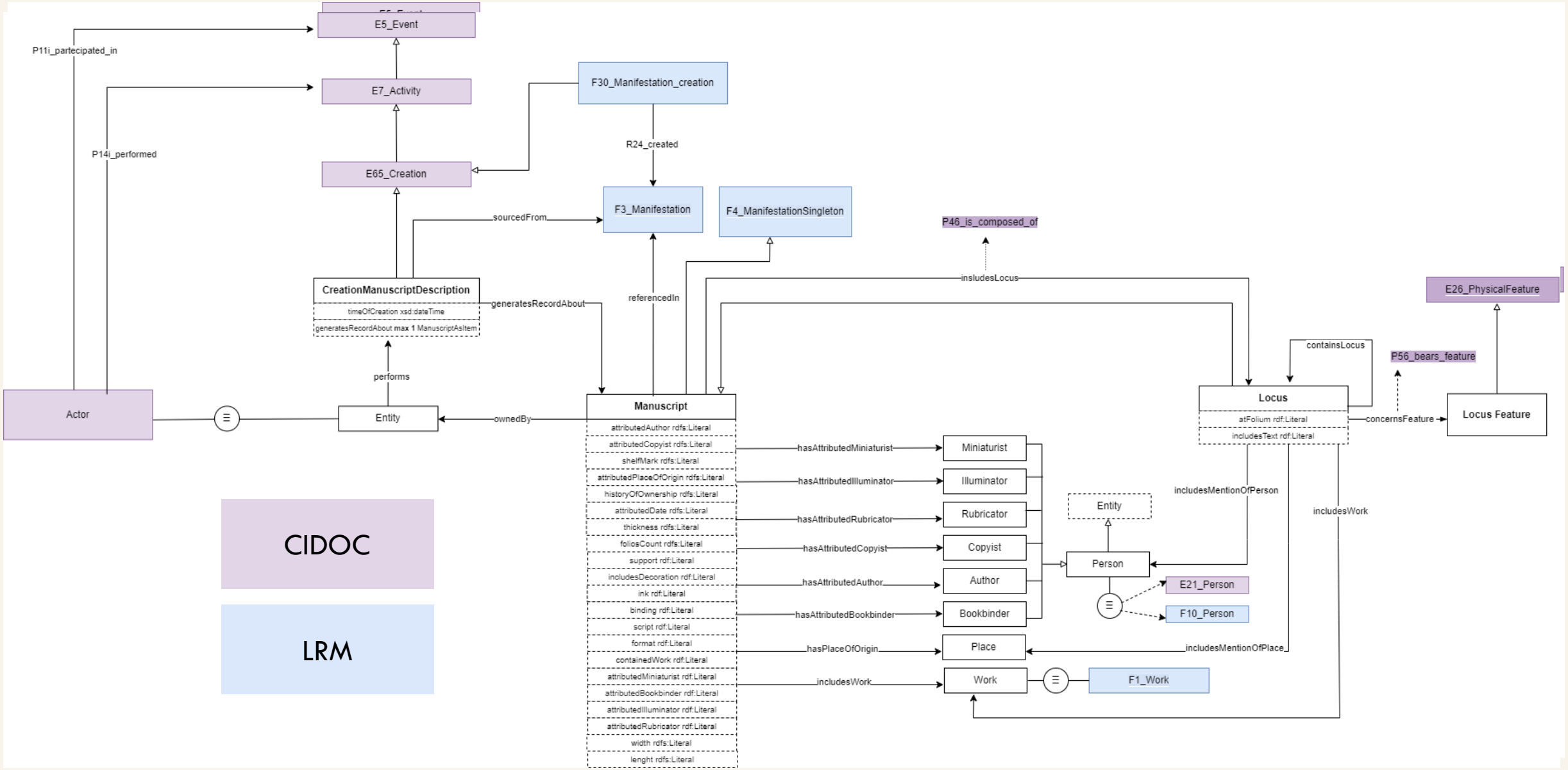


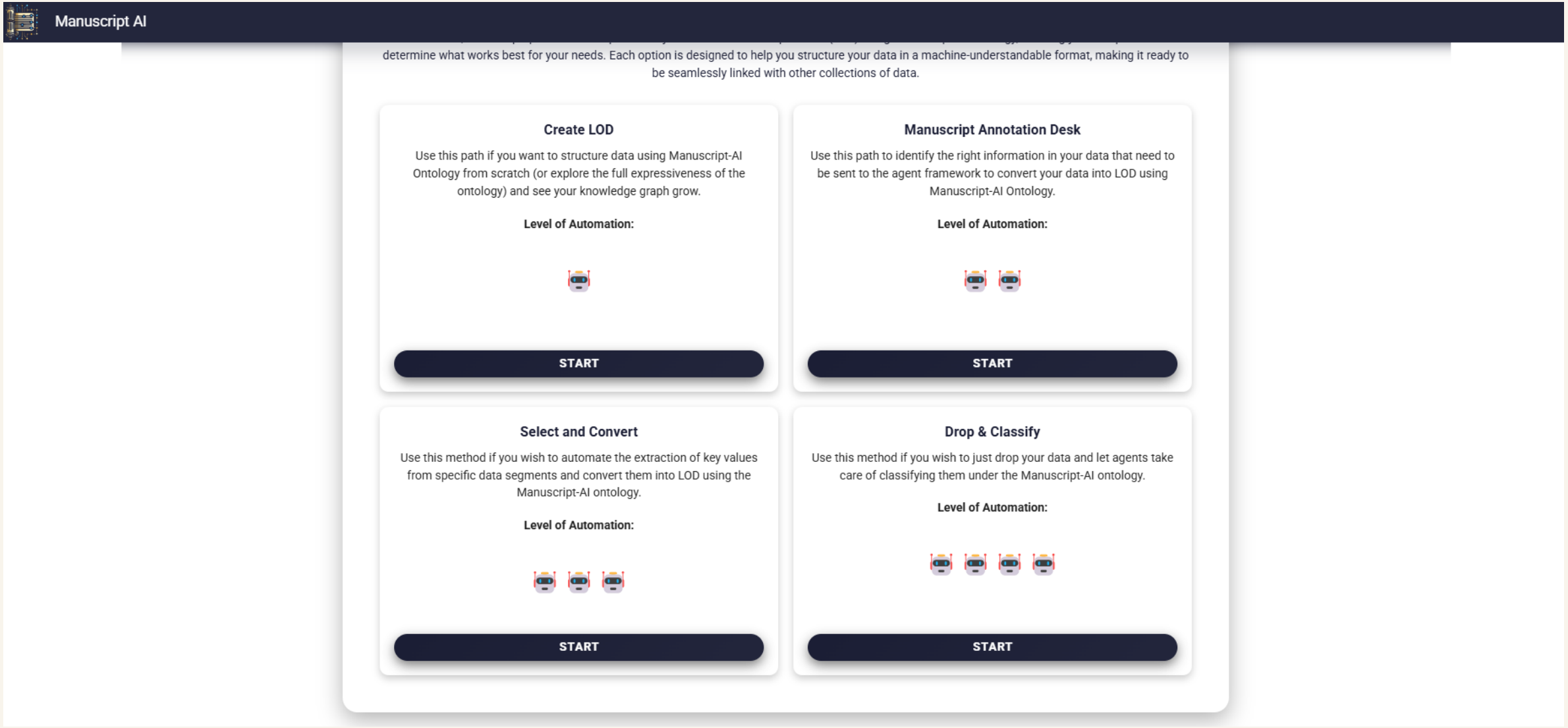
Evaluation & Pilot

- After the conceptual design and development of the Manuscript AI ontology and Agentic Framework: **evaluation and pilot deployment**
- Test the system **in a real-world heritage context**, engaging domain experts and heritage institutions.
- The pilot outcomes feed back directly into refining the ontology: A) Expanding coverage where needed; B) Adjusting formal mappings or alignments to CIDOC-CRM to ensure consistency and interoperability.

In short: **design → test → refine** is the chosen cycle, enabling Manuscript AI to evolve based on concrete use.


F3_Manifestation
sourcedFrom
referencedIn
includesLocus
CreationManuscriptDescription
timeOfCreation xsd:dateTime
generatesRecordAbout max 1 ManuscriptAsItem
generatesRecordAbout
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
foliosCount 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
Locus
atFolium rdf:Literal
includesText rdf:Literal
containsLocus
concerns
Locus Feature
illumination
decoratedInitial
schematicDrawing
miniature
penworkInitial
coloredDrawing
historiedInitial
figure
ornamentation
borderDesign
illustrationCycle
panel
drollerie
illustration
tooling
clasp
fastener
embossing
rdf:type
Decoration
hasDecoration
copticBinding
carolingianBinding
romanesqueBinding
gothicBinding
limpVellumBinding
sewnOnCordsBinding
BindingTechnique
hasBIndingTechnique
quarto(4to)
duodecimo(12mo)
folio(2to)
sextocedimo(16mo)
octavo(8vo)
Format
hasFormat
Support
hasSupport
papyrus
donkeySkin
marbledPaper
parchment
uterineVellum
russiaLeather
silk
palmLeaves
calico
bark
paper
canvas
vellum
sheepskin
velvet
naturalGoatskin
roughskin
satin
deerSkin
pigskin
morocco
hasAttributedMiniaturist
hasAttributedIlluminator
hasAttributedRubricator
hasAttributedCopyist
hasAttributedAuthor
hasAttributedBookbinder
hasPlaceOfOrigin
includesWork
hasInk
Miniaturist
Illuminator
Rubricator
Copyist
Author
Bookbinder
Place
Work
F1_Work
Entity
Person
includesMentionOfPerson
includesMentionOfPlace
includesWork
Ink
carbonInk
invisibleInk
copperGallInk
redInk
ironGallInk
coloredInk
OrganicInk
hasScript
Script
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
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


E5_Event
P11i_partecipated_in
E7_Activity
F30_Manifestation_creation
P14i_performed
R24_created
E65_Creation
F3_Manifestation
F4_ManifestationSingleton
sourcedFrom
P46_is_composed_of
insludesLocus
CreationManuscriptDescription
timeOfCreation xsd:dateTime
generatesRecordAbout max 1 ManuscriptAsItem
generatesRecordAbout
referencedIn
E26_PhysicalFeature
containsLocus
P56_bears_feature
performs
Actor
Entity
ownedBy
Manuscript
attributedAuthor rdfs:Literal
attributedCopyist rdfs:Literal
shelfMark rdfs:Literal
attributedPlaceOfOrigin rdfs:Literal
historyOfOwnership rdfs:Literal
attributedDate rdfs:Literal
thickness rdfs:Literal
foliosCount rdfs:Literal
support rdf:Literal
includesDecoration rdf:Literal
ink rdf:Literal
binding rdf:Literal
script rdf:Literal
format rdf:Literal
containedWork rdf:Literal
attributedMiniaturist rdf:Literal
attributedBookbinder rdf:Literal
attributedIlluminator rdf:Literal
attributedRubricator rdf:Literal
width rdfs:Literal
lenght rdfs:Literal
hasAttributedMiniaturist
hasAttributedIlluminator
hasAttributedRubricator
hasAttributedCopyist
hasAttributedAuthor
hasAttributedBookbinder
hasPlaceOfOrigin
includesWork
Miniaturist
Illuminator
Rubricator
Copyist
Author
Bookbinder
Place
Work
F1_Work
Locus
atFolium rdf:Literal
includesText rdf:Literal
concernsFeature
Locus Feature
includesMentionOfPerson
includesWork
Entity
Person
E21_Person
F10_Person
includesMentionOfPlace
CIDOC
LRM


Manuscript AI
determine what works best for your needs. Each option is designed to help you structure your data in a machine-understandable format, making it ready to be seamlessly linked with other collections of data.
Create LOD
Use this path if you want to structure data using Manuscript-AI Ontology from scratch (or explore the full expressiveness of the ontology) and see your knowledge graph grow.
Level of Automation:
START
Manuscript Annotation Desk
Use this path to identify the right information in your data that need to be sent to the agent framework to convert your data into LOD using Manuscript-AI Ontology.
Level of Automation:
START
Select and Convert
Use this method if you wish to automate the extraction of key values from specific data segments and convert them into LOD using the Manuscript-AI ontology.
Level of Automation:
START
Drop & Classify
Use this method if you wish to just drop your data and let agents take care of classifying them under the Manuscript-AI ontology.
Level of Automation:
START

Create LOD path: Data Insertion & Real-Time Growth:

- The user manually inputs initial manuscript data.
- As each new value is added, the knowledge graph (KG) immediately expands, allowing the user to see real-time changes.

Manuscript-AI Annotation Desk path

- The user works within the Manuscript-AI annotation desk to assign specific values (e.g., attributed author, century).
- The agent framework converts this annotated data into RDF, aligned with the Manuscript-AI ontology, and enriches it with links from Wikidata.

Select and Convert Path

- The user selects only a portion of the data (chunk) for a specific manuscript.
- The agent framework automatically identifies and proposes relevant values for key categories.
- The user reviews and corrects these suggestions if needed.
- The refined data is then converted into RDF (Manuscript-AI ontology) and enriched with external references.

Drop and Classify Path: Automated Chunking & Enrichment

- The chunking of data happen automatically.
- The user's role is primarily to validate final results
- The agent framework structures the data, applies the Manuscript-AI ontology, and provides enrichment links, with only final confirmation or adjustments required from the user.

The pilot phase began on **August 1, 2025**, and is currently ongoing

25 participants from **13 different institutions**

Evaluation Goals:

- Assess AI structuring of catalogue data
- Evaluate ontology coverage
- Measure interface usability

Two Feedback Instruments: A) Structured questionnaire (pre- and post-test) and **B)** Ontology feedback spreadsheet

Questionnaire Process

Pre-test: Participants report background and experience with manuscript-related digital tools

Post-test:

- Rate interface usability (Likert scale)
- Open questions on preferred processing paths, missing features, bulk processing
- Self-report workload (effort, time, frustration)

Ontology Feedback

Spreadsheet to log missing classes/properties, and modelling gaps

The pilot revealed **valuable insights** into areas where **Manuscript AI Ontology** can further evolve to capture complex manuscript realities.

Participants identified **rich opportunities for expansion** in representing:

Palimpsests : modeling the layered nature of writing and textual reuse.
Composite manuscripts: expressing the combination of multiple origins and provenances.
Disiecta membra : describing the relationships among dispersed or recombined fragments.

These cases open avenues to **strengthen the ontology's capacity** for representing codicological phenomena.

The analysis of the **coverage and expressivity feedback** from the pilot indicates that the Manuscript AI ontology achieves a **well-considered balance** between the **granularity required for manuscript description** and the **broader interoperability** needed for cultural heritage integration.

Through the analysis of the **expressivity mapping results**, it became clear that Manuscript AI and **CIDOC-CRM** operate in a **complementary relationship.**

While Manuscript AI focuses on the **specialized semantics** of manuscript studies, CIDOC-CRM provides the **broader conceptual structures** needed to place those descriptions within historical, institutional, and further event-based contexts.

This complementarity establishes a **continuum of representation**, ensuring that detailed manuscript data can be connected to higher-level heritage information without losing conceptual accuracy or context.

Feedback from the pilot highlighted the **importance of representing degrees of certainty** and **levels of confidence** in descriptive assertions. This need is particularly relevant for aspects such as **authorship attribution**, **dating hypotheses**, or the **identification of scriptoria**, where information often remains interpretative or uncertain.

Introducing mechanisms for **uncertainty modeling** would:

- Allow a more nuanced expression of **hypotheses, conflicting evidence, and provisional knowledge**.
- Support **comparative reasoning** across alternative interpretations or scholarly viewpoints.
- Promote **transparency and accountability** in the way digital manuscript descriptions record interpretive judgments.

Future development could explore **formal strategies** (such as confidence indicators, provenance annotations, or probabilistic relationships) to capture this **epistemic dimension** and reflect the interpretive nature of manuscript scholarship.

FUTURE ACTIONS

Integrating pilot feedback:

- The next phase will focus on incorporating the **input and insights from the pilot evaluation** to refine the ontology's coverage and expressivity
- Particular attention will be given to enhancing the modeling of **codicological complexity**, **transformations**, and **contextual relationships** identified during the pilot.

Formalization and technology-agnostic representation:

- The ontology will be **translated into a formal, machine-readable representation** to preserve conceptual clarity and ensure **technology independence.**
- This process will also consolidate **alignment with CIDOC-CRM**, strengthening interoperability across cultural heritage data models.

FUTURE ACTIONS

Towards integration within CIDOC-CRM-compatible modules:

The ongoing work aims to ensure that the Manuscript AI conceptualization remains **consistent with CIDOC-CRM principles** and extensible through its **modular architecture**.

The goal is to make the ontology **ready for potential inclusion or linkage** within the **CIDOC-CRM compatible family of models**, where it could contribute a manuscript-focused, fine-grained layer of representation.

Collaboration with the CIDOC-CRM community will be essential to determine **the most appropriate integration pathway** and to ensure conceptual coherence with existing modules.

THANK YOU

S. Boodts & G. Biagioni

shari.boodts@ru.nl & giulia.biagioni@tno.nl

ManuscriptAI | Radboud Universiteit (ru.nl)

European Research Council
Established by the European Commission