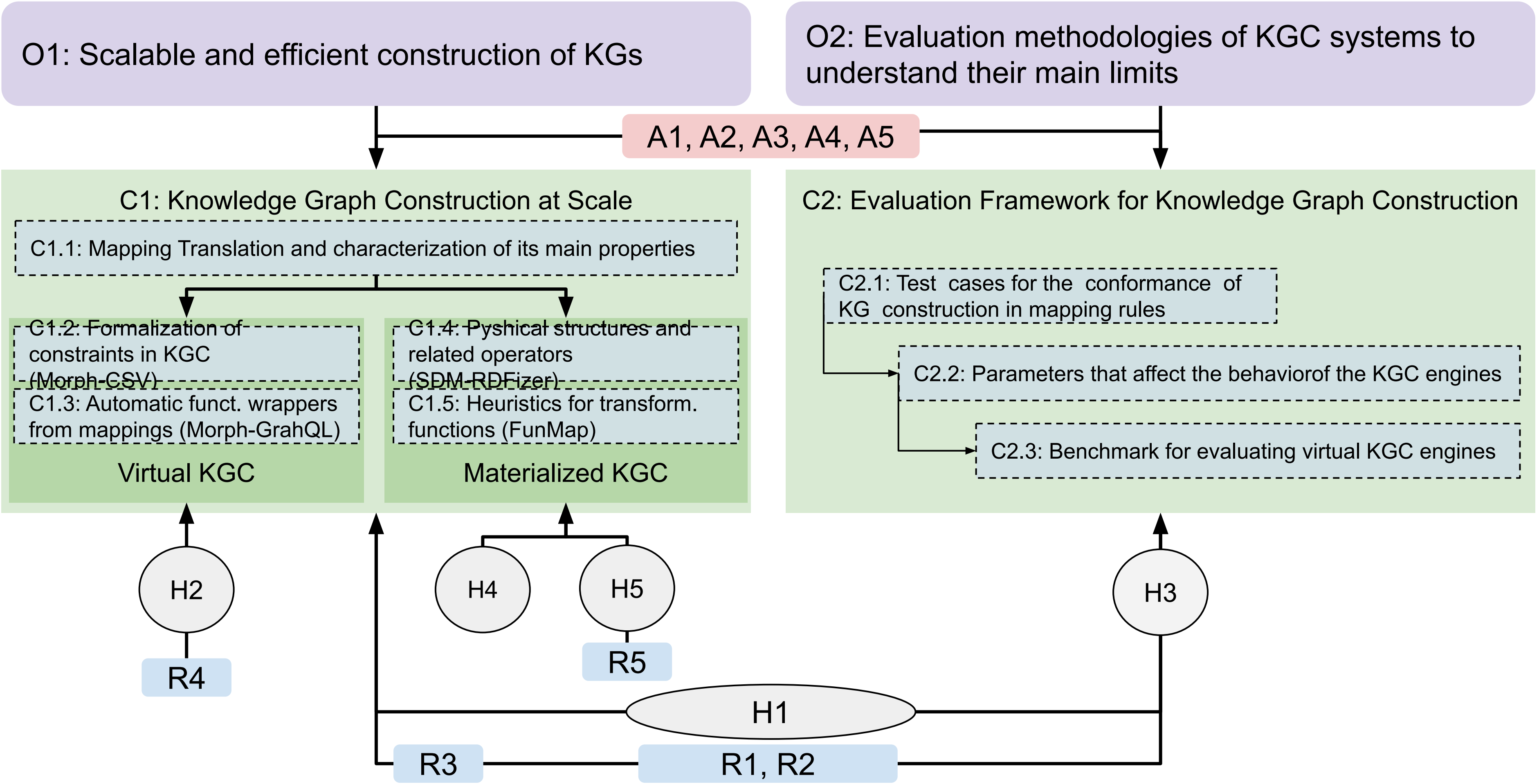
New Generation of Knowledge Graph Construction Engines



O1: Scalable and efficient construction of KGs

O2: Evaluation methodologies of KGC systems to understand their main limits

A1, A2, A3, A4, A5

C1: Knowledge Graph Construction at Scale

C1.1: Mapping Translation and characterization of its main properties

C1.2: Formalization of constraints in KGC (Morph-CSV)

C1.3: Automatic funct. wrappers from mappings (Morph-GrahQL)

Virtual KGC

C1.4: Pyshical structures and related operators (SDM-RDFizer)

C1.5: Heuristics for transform. functions (FunMap)

Materialized KGC

C2: Evaluation Framework for Knowledge Graph Construction

C2.1: Test cases for the conformance of KG construction in mapping rules

C2.2: Parameters that affect the behaviorof the KGC engines

C2.3: Benchmark for evaluating virtual KGC engines

H2

R4

H4

H5

R5

H3

H1

R3

R1, R2

Assumptions:

A1: Mapping rules and metadata descriptions are declarative and follow W3C standards
A2: The ontology for integrating the source data is available and is implemented in OWL.
A3: Mapping rules and metadata are available
A4: Data are represented in formats that are not RDF
A5: Datasets are static, not streams.

Restrictions:

R1: Input data sources must be located in the same physical place as the KGC process is run.
R2: Not have to consider data protection nor access restrictions.
R3: The size of datasets is defined in terms of Gigabytes
R4: Our proposal does not make use of the capabilities of SPARQL-to-SQL engines
R5: Our proposal does not make use of the features of the transformation functions performed over the input data sources.

Hypotheses:

H1: It is possible to translate declarative mapping rules among different specifications.
H2: The exploitation of declarative annotations can enhance current virtual KGC systems
H3: A benchmark on transport data is able to stress and provide a full overview of the current state of different KGC engines
H4: Physical data structures and operators can be defined for scaling up KGC engines
H5: Optimizations for functional mapping rules can be applied to scale up the construction of KGC