IRTF NMRG · IETF 127

Portable Semantic Models and Ad Hoc Reconciliation by Cognitive Agents in Network Management

draft-janz-nmrg-adhoc-semantic-reconciliation-01

Chris Janz, Huawei · Brad Peters, NBN

THE QUESTION

Do systems that can reason still need a model agreed in advance?

Agreed in advance: today

A data model, or a richer information model or ontology, is agreed before two systems interoperate.

That takes broad prior agreement, negotiated and maintained by people.

Reconciled ad hoc: this draft

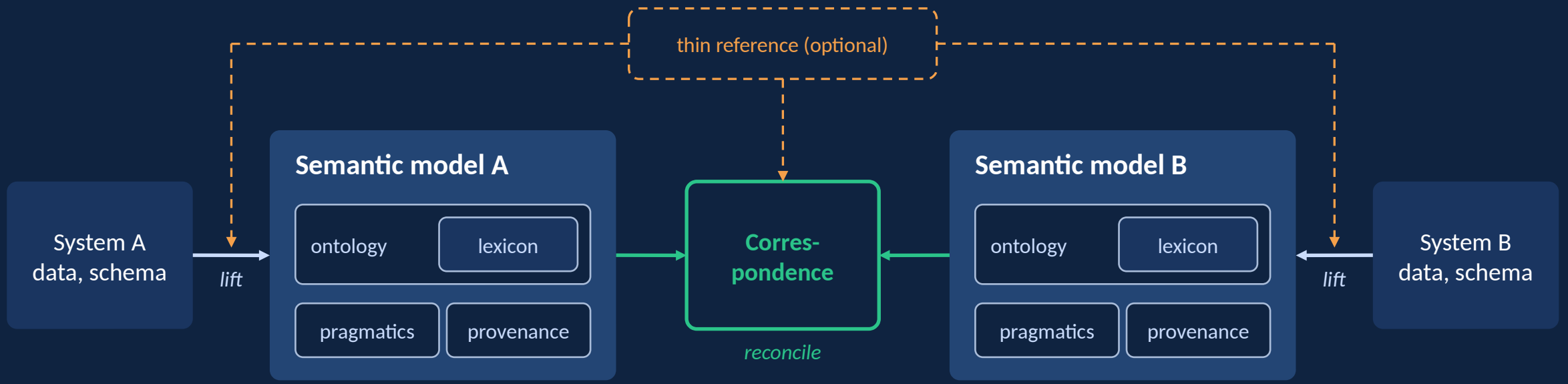
Each side lifts its own data into a complete semantic model.

Two such models reconcile for the occasion, with no model agreed beforehand.

The study's answer: capable agents can build what they need, within limits this talk states.

THE ACCOUNT IN ONE PICTURE

Two operations: lift and reconciliation



Ontology
concepts, relations and instances

Lexicon
names, synonyms and definitions

Pragmatics
what a thing is for, whose judgment governs, what a signal means in context

Provenance
who asserted a fact, by what method, how firmly

Portable: another agent can use the model from the package alone.

Lift and reconciliation each run as a **reasoning-and-acting loop**.

HOW THE STUDY MEASURED

Four scenarios, three model families, authored answer keys

Four scenarios, each on three independently built cases

Configuration	ONF TAPI against an IETF TE topology
Intent	A TM Forum TR290 intent into a layer-1 realization
Standard-free	No standard to lean on
Observability	Alarm against anomaly, RFC 9940

Agents from three families: OpenAI GPT at a strong, a middling and a weak tier; DeepSeek; the open model Qwen.

Measuring the lift

Comprehension score

Portability: the share of a model's concepts an independent agent, given only the package, explains faithfully, as graded by a fixed judge against the key

Measuring the reconciliation

Resolved fraction

The share of the true correspondences an agent finds and commits

Precision

The share of what it commits that is correct

Reconciliation effort is counted in reasoning tokens.

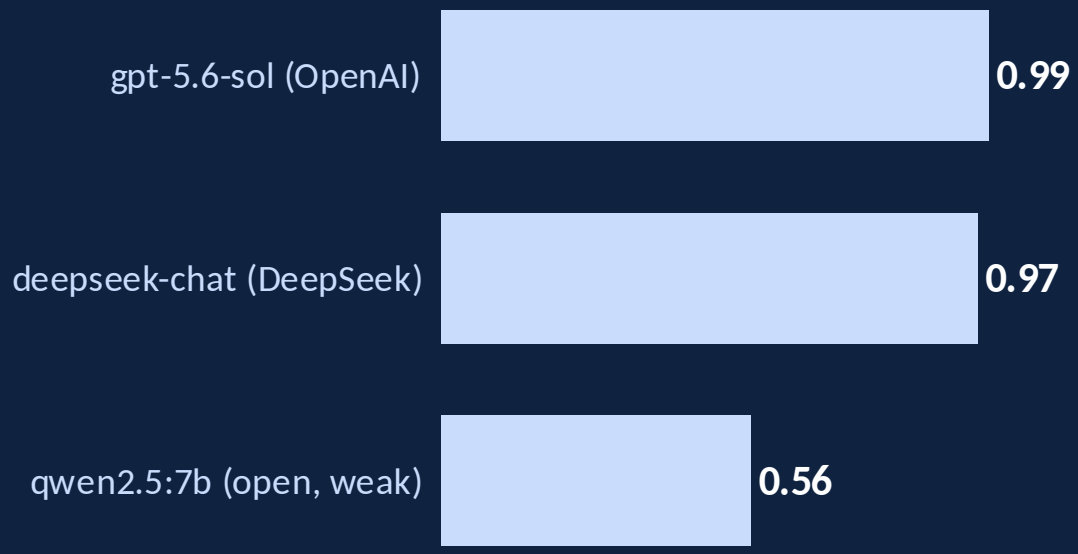
FINDING 1 · PORTABILITY

A well-lifted model is understood cold, from the package alone

One family, three tiers
the hardest material: concepts from two unrelated domains



Other families, one cross-family test
a separate test; GPT's strong tier was run on it too, for comparison

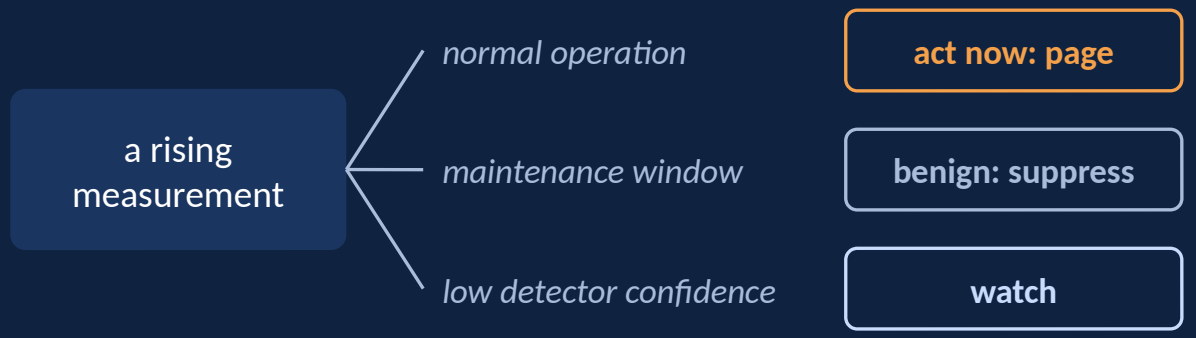


Comprehension score: the share of a model's concepts a reader, given only the package, explains faithfully.

A soft gradient, not a cliff: capability sets the score, not family. Qwen, bimodal and confabulating, is the floor.

Pragmatics carries the verdict, and a schema rarely holds it

One signal, three contexts, three verdicts



What the agent was given, for one anomaly

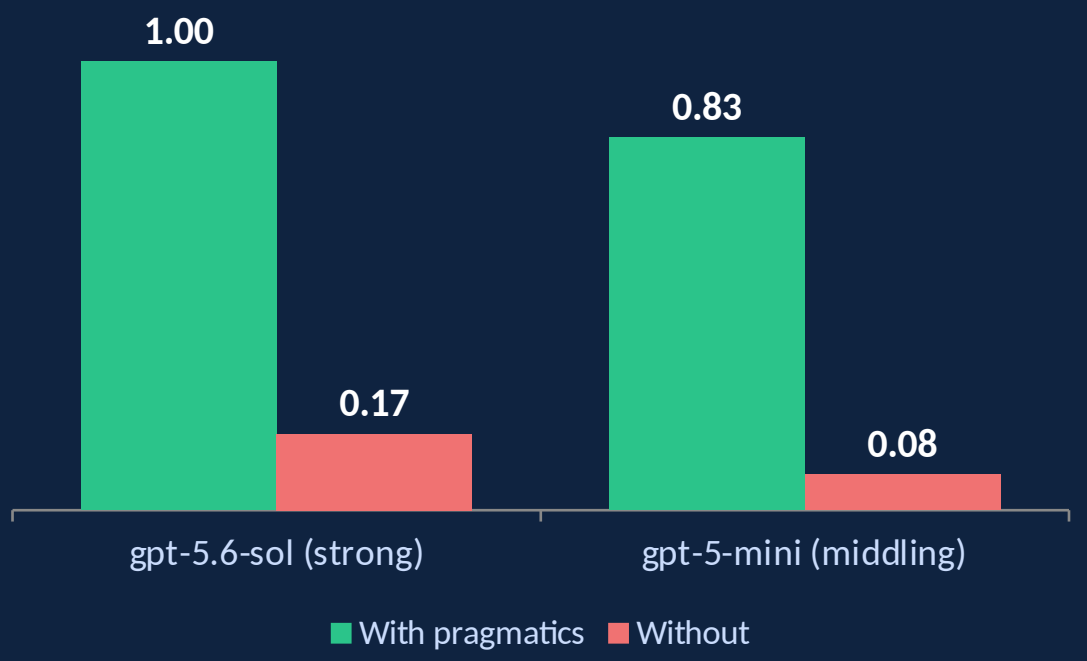
With pragmatics

```
A1 "rising pre-FEC BER on lambda1"
concern 78, confidence 82 (of 100)
plane forwarding, pattern trend,
lifecycle validation, not seasonal
context: no maintenance, a workday
```

Without

```
A1 "rising pre-FEC
BER on lambda1"
severity critical
no scores, no
context
```

Significance verdict accuracy, observability



Without pragmatics, the same agents paged nearly everything.

Reconciliation refuses a look-alike rather than binding it

Transport “grade”

A protection class

≠

IP “grade”

A class of service

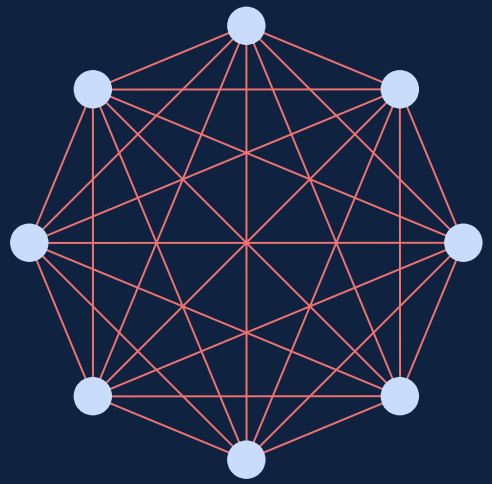
A label matcher binds them, because the strings match. A capable agent declines the pairing on evidence that is not lexical: the concepts' kinds, what they attach to, and the instances they hold.

Where it cannot ground a pairing, it defers rather than guesses. That is what protects precision.

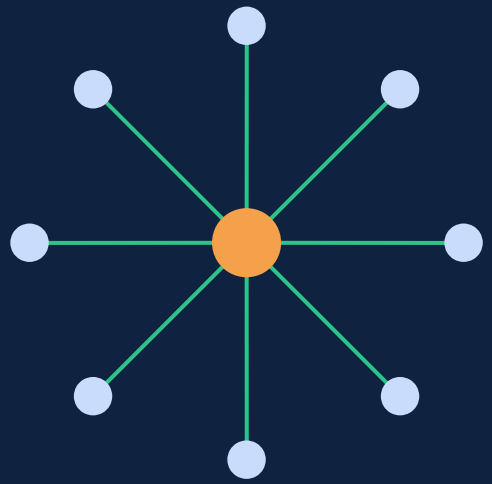
A thin reference does two jobs, and binding to it scales linearly

What it is and what it does

- Thin: an identifier, a label and synonyms, a shallow class, a definition and one example. A lexicon, not a domain model.
- Two roles: it supplies meaning to one consumer, and common ground to a pair.



pairwise: N(N−1)/2



via one reference: N

Checked out to twelve systems: composing the single-reference bindings reproduces the correct pairwise correspondences.

Caution: the description binds, never a bare pointer. An identifier with no description is worse than no reference.

Where no standard exists, agents build the common ground

Resolved fraction, standard-free scenario, strong agent

0.20

No reference

It refuses to guess the seam between the two models; every commitment is correct

→

0.80

Agents build a thin reference

Built from the two models, then bound through

→

0.90

Plus one virtual experiment

Provision a candidate correspondence on the live side and read back whether it holds

Precision stays at 1.00 throughout. The climb is in what agents surface, not a trade against correctness.

THE COGNITION SPECTRUM

Where cognition sits decides what a reconciliation can guarantee

Inert: a system with no reasoning of its own, such as a legacy system or a frozen export. Sound: no false correspondence. Complete: every true one resolved.

cognition recedes →

Both sides cognitive

Means: questions to a live counterpart, virtual experiment, record check.

Sound and complete: reconciliation always completes.

about 16×

fewer reasoning tokens when a strong agent is given a reference

One side inert

Means: questions and experiment on the live side, record check, reference.

Sound; completeness not guaranteed.

about 6×

fewer reasoning tokens when a strong agent is given a reference

Both sides inert

Means: the record check alone.

Sound; lower coverage, no decisive experiment.

A reference recovers coverage the record check misses.

Full autonomy holds at the cognitive end. Elsewhere a thin reference buys back part of what is lost.

EVIDENCE AND A CAUTION

The results hold, with one failure mode to plan for

Configuration

ONF TAPI against an IETF TE topology. Precision and resolved fraction 1.00.

Intent

A TMF TR290 intent negotiated into a layer-1 realization. Decision accuracy 1.00.

Standard-free

No standard to lean on. Resolved fraction 0.20 to 0.90 at precision 1.00.

Observability

RFC 9940 alarm against anomaly. Verdict accuracy 1.00 and 0.83 with pragmatics.

The benchmark is released, so every result can be rechecked: doi:10.5281/zenodo.23187404.

Confident errors rise about elevenfold down the capability ladder, from 0.004 to 0.046

Confidence on wrong bindings sat at 0.73 to 0.85, close to that on correct ones, so a confidence cutoff cannot separate them. Weak or inert-facing reconciliation needs an explicit abstain path.

IMPLICATIONS FOR STANDARDIZATION

What is worth agreeing in advance narrows, and changes kind

Thin references

An identifier, a label and synonyms, a shallow class, a definition and one example: a lexicon, not a domain model. Cheap to maintain and broadly reusable.

Authority boundaries

Who governs a contested field. No reasoning settles it: it is governance, not modelling, and standardization stays indispensable here.

Self-description and interrogation

How a system exports its surface, explains its concepts, declares authority and provenance, and answers open questions.

“Agree how a system explains itself, not the model it uses inside.” An evolution, not a rupture.

NEXT STEPS AND ASKS

Next: re-measure on operator data, then lift a live system in place

1 Operator data at production scale

Re-establish portability, resolved fraction, precision, the reference's value and the effort cost on real, messy models.

2 A situated lift on a live management plane

A capable agent with read-only tools over a TAPI controller, a YANG datastore over NETCONF or RESTCONF, or a gNMI stream; budgeted queries, rate limits, full audit, no write path. Sweeping from frozen export to self-lift separates what a lift can recover from what is a matter of authority.

Asks of NMRG and NMOP

- Operators: run the situated lift yourselves, with a capable agent and read-only access to the datastores that suit the case; the draft gives the whole portability assessment method (Section 4.2)
- Report the value of the lift: portability of the lifted model minus that of the un-lifted baseline
- Views on the three candidates for standardization
- Links with NMOP knowledge-graph and AI-modelling work
- Comments on the draft

TAKEAWAYS

Systems that can reason can carry their meaning with them

1

Portable models are real

Capable readers understand a well-lifted model cold. Pragmatics has to be in it.

2

Reconciliation can be ad hoc

Disciplined and verified, it builds missing ground and always completes when both sides reason.

3

Supporting standards evolve

To thin references, authority boundaries and self-description (API).

Questions and comments

datatracker.ietf.org/doc/draft-janz-nmrg-adhoc-semantic-reconciliation/ · benchmark doi:10.5281/zenodo.23187404