Online Resource 1: Supplementary Material

for

Capability-based domain assignment for patent-based convergence analysis in classification-dispersed fields: an ISRU space-construction case

Scientometrics

Tai Sik Lee[a,b], Yoonsun Lee[c,*]

[a] E&I Korea, Seoul 04921, Republic of Korea

[b] Civil and Environmental Engineering, Hanyang University, Ansan 15588, Republic of Korea

[c] Research Institute of Engineering Technology, Hanyang University, Ansan 15588, Republic of Korea

* Corresponding author. E-mail address: yoonsunlee@hanyang.ac.kr (Y. Lee)

This Online Resource provides supporting tables, accounting details, and notes for the main manuscript. Values reported here are aligned with the released analytical package. In this study, classification-dispersed refers to dispersion across CPC prefixes and title/abstract vocabularies within a common patent-classification infrastructure, not incompatibility among national classification systems.

Supplementary map. Guide to the supplementary items and their roles.

Supplementary item	Main role
S1, S3	Background external architecture and WBS plausibility context
S2	Jurisdiction-stratified sensitivity
S4, S8, S12	Adjacency interpretation and CPC-anchor sensitivity
S5, S7-S7c	Domain-external accounting and extension priorities
S6, S6a, S9	Rule-hit-to-assignment diagnostics
S10	CPC bridging centrality
S11	Reconstruction scope of the released package
S13	Rescue-route sensitivity
S14	Deterministic first-pass and WBS–ITC assignment comparison
S15	Comparison materials note
S16	Expanded deterministic and restrictive assignment-variant comparison
S17	Supported uses, scope boundaries, and additional analyses needed for broader claims

Table S1. Illustrative ISRU-relevant and analog non-patent exemplars mapped to WBS–ITC domains.

Project or example	Documentation basis	Stage	Indicative WBS–ITC mapping	Why included here
Lunar Contour Crafting	Khoshnevis et al. (2005)	Concept / early system proposal	2-1; 4-1	Canonical construction-printing analog for lunar habitat fabrication.
Building components from lunar soil by 3D printing	Cesaretti et al. (2014)	Prototype / component demonstration	1-3; 2-1; 4-1	Illustrates regolith-derived structural fabrication.
Automated Additive Construction (AAC)	Mueller et al. (2016)	Demonstration platform	2-1; 2-4; 3-1	Useful non-patent analog for AM plus robotic execution.
Extra-terrestrial construction process review	Lim et al. (2017)	Synthesis / process framing	1-3; 2-1; 3-1; 4-1	Supports staged construction-chain reading used in the manuscript.
Solar 3D printing of lunar regolith	Meurisse et al. (2018)	Laboratory demonstration	1-3; 2-3	Energy-based consolidation pathway for regolith-derived structures.
Vacuum sintering of lunar regolith simulant	Reiss et al. (2019)	Laboratory demonstration	1-3; 2-3	Material-process plausibility check under lunar-like constraints.
Lunar and Martian regolith-based concrete as building blocks for future human settlements	Heemskerk et al. (2020)	Conference abstract / materials study	1-2; 1-3	Supports WBS-1 materials layer mapping logic for regolith-based binder/composite formulations.

Project or example	Documentation basis	Stage	Indicative WBS–ITC mapping	Why included here
RoboCrane lunar cave support system	Miaja et al. (2022)	Subsystem demonstration	3-1; 3-2; 4-4	Illustrative robotics / deployable-support analog within the broader construction chain.
Lunar base construction overview	Mueller (2022)	Programmatic capability overview	2-1; 3-1; 4-1	Supports field-boundary construction from external architecture rather than patent classes.
KICT lunar habitation core technology (phase 1)	KICT (2023)	R&D program / taxonomy	1-2; 1-3; 2-1; 4-1	Construction-oriented external taxonomy used in WBS synthesis.
KICT lunar habitation core technology (phase 2)	KICT (2024)	R&D program / taxonomy	1-2; 2-1; 4-1; 4-4	Provides additional construction-process framing for the specification.
Autonomous robotics lunar landing-pad construction simulation	Okonkwo et al. (2026)	Task-level simulation	3-1; 3-3; 4-3	Useful plausibility check for execution-layer extension around robotic construction tasks.
Large-scale ISRU and surface construction infrastructure architecture	Sanders et al. (2024)	NASA architecture / concept-of-operations framing	WBS-1; WBS-2; WBS-3; WBS-4	Frames ISRU construction as a long-term lunar infrastructure capability chain linking resource processing, construction feedstocks, autonomous operations, power, communications/PNT, maintenance, and surface civil engineering.

Note. These exemplars are used as plausibility checks for the WBS–ITC logic and as external context for capability-architecture construction; they do not enter the empirical calculations.

Table S2. Jurisdiction-stratified sensitivity analysis.

Panel A. Dataset scope and jurisdiction composition.

Item	Released set	Active set	Interpretive note
Total families	453	443	Ten records are excluded from reported analytical modules (9 false matches + 1 documentation-only below-threshold record).
CN families	342 (75.5%)	336	Chinese-jurisdiction filings dominate the analytical layer numerically. CN/non-CN status is assigned by earliest priority jurisdiction as recorded in the Lens family metadata. Panel A reports family counts under this criterion, and Panel B reports WBS-layer tag shares within the corresponding classified subsets; percentages are calculated over final-assignment ITC tags within each jurisdictional classified subset.
Non-CN families	111 (24.5%)	107	Non-CN subset remains analytically meaningful for sensitivity checking.
ITC-classified families	373	373	ITC-based tag shares and Jaccard values are invariant to the 10 exclusions because excluded cases carry no ITC tags.

Panel B. WBS-layer tag shares among classified families.

WBS layer	All	CN	Non-CN	Interpretation
Materials	28.1%	29.0%	25.5%	Materials remain substantial in both subsets.
Manufacturing	31.8%	35.8%	21.3%	Manufacturing emphasis is stronger in the CN subset.
Robotics	4.7%	4.3%	5.6%	Robotics remains a narrower claim-supported layer across subsets.
Structures & Systems	35.5%	30.9%	47.7%	Structures become relatively more prominent in the non-CN subset.

Panel C. Stability of leading adjacency and bridging patterns.

Analytical check	All families	CN subset	Non-CN subset	Interpretation
Top Jaccard pair	2-1 <-> 2-4 (J = 0.341)	2-1 <-> 2-4 (J = 0.333)	2-1 <-> 2-4 (J = 0.381)	The leading overlap pair persists across both subsets.

Analytical check	All families	CN subset	Non-CN subset	Interpretation
Top bridging code	B64G	G01N	B64G	Principal bridging position remains space/measurement related even though exact rank order shifts.

Note. Interpretively, jurisdiction affects relative emphasis, but the core WBS–ITC assignment structure remains recognizable across the split samples.

Table S3. Schematic cross-agency priority framing used to motivate the WBS architecture.

Agency / source family	Programmatic framing	Construction-related implication	Role in present study
ISECG	Long-horizon exploration architecture; staged ISRU and civil-engineering logic	Supports reading lunar construction as a capability chain rather than a single mission class.	Used as a major external input for four-layer WBS synthesis.
NASA	Moon-to-Mars architecture and AECO / Swamp Works construction framing	Provides construction-process and discipline-level decomposition relevant to fabrication, emplacement, and habitats.	Supports WBS-2, WBS-3, and WBS-4 articulation.
ESA	Space-resources value chain and lunar infrastructure enablers	Places construction downstream of resource-access and logistics functions.	Used mainly to justify transfer-inclusive boundary setting and downstream structural interpretation.
JAXA	Water-resource exploration near term; plant capability later	Reinforces staged progression from prospecting to processing and utilization.	Supports scope note that the present specification captures a bounded construction segment, not the whole exploration chain.
CNSA	South-polar resource prospecting linked to later utilization experiments	Shows programmatic coupling between reconnaissance and subsequent resource-use architectures.	Used as context rather than as a direct determinant of ITC domain boundaries.
KICT	Construction-oriented 6-19-87 technology tree	Provides construction-language decomposition that is closer to engineering tasks than patent classes are.	Key source for translating agency architecture into boundary specification domains.

Note. This table provides interpretive context only. It situates the WBS–ITC instrument within broader external mission and construction logics and does not enter the empirical calculations.

Table S4. Co-tagging interpretation summary.

Adjacency pair	Evidence in analytical layer	Co-tagging interpretation	Interpretive use
2-1 <-> 2-4	Highest Jaccard pair (0.341); 46 co-tagged families; 89% with earliest priority 2020 or later	Frequent co-tagging within manufacturing domains, partly conditioned by shared classification structure	Candidate adjacency zone around additive manufacturing and process monitoring
4-1 <-> 4-4	High overlap (0.294); 48 co-tagged families	Attenuated but persistent structural co-occurrence within habitat-related domains	Persistent adjacency signal within the structural layer under the tested anchor removal
1-3 <-> 2-1	Moderate overlap (0.172)	Co-tagging between material formulation and extrusion processes	Recurrent co-tagged material-process overlap within WBS–ITC assignment
1-1 <-> 2-4	Top-10 pair (0.167)	Co-occurrence between processing and monitoring domains	Candidate material-process adjacency within the released WBS–ITC assignment
WBS-3 <-> WBS-4	Robotics accounts for 4.7% of total tags; no robotics-structures pair appears in the top-ranked Jaccard list	Narrower claim-supported execution-layer co-tagging within the current specification	Boundary condition of the present framework
4-2 <-> 4-5	Top-10 pair (0.163)	Co-tagging within protection and life-support domains	Candidate co-tagging zone within protection and life-support domains

Note. Entries summarize co-tagging patterns discussed in the main text.

Table S5. Domain-external family disposition and analytical treatment.

Disposition class	N	Analytical status	Interpretive role
Active domain-external families	70	Retained in the 443-family active set	Construction-relevant but not assigned into the current 15-domain specification; used as boundary evidence for extension.
False matches	9	Excluded from all reported analytical modules	No plausible construction-relevance pathway to ISRU construction under the study construction-relevance criteria.
Documentation-only below-threshold record	1	Excluded from all reported analytical modules	Retained only for documentation of construction-relevance logic, not for reported analysis.
Total non-ITC-classified records in released dataset	80	Visible in the 453-family release	Combines active domain-external cases and excluded non-ITC-classified records for full accounting transparency.

Note. The accounting logic is reported explicitly because the manuscript relies on the distinction between released families, active set, and ITC-classified subset.

Table S6. Rule-hit layer and WBS–ITC assignment structure.

Comparison item	Rule-hit candidate generation	WBS–ITC assignment	Why the distinction matters
Primary inputs	Published CPC anchor prefixes and title/abstract keyword stems (OR logic)	Representative abstract and first independent claim read against the ITC domain definitions	Separates transparent candidate generation from capability-fit assessment.
What a positive signal means	A family is flagged as a candidate for a domain	A family receives one or more WBS–ITC tags	The rule set is not treated as the final classification object.
Permitted assignment changes	Not applicable	May add domains with no rule hit or remove rule-positive domain hits after claim-based capability-fit assessment	This is where claim-supported domain fit is evaluated.
Layer used for reported tables and figures	No	Yes	All manuscript ITC results are computed from the assignment set.
Reconstruction role	Documented first-pass rule structure	Released assignment dataset and scripts for downstream reconstruction	Reconstruction is defined as rebuilding reported outputs from the released assignment dataset and scripts.
Implication for reported analysis	Supports transparency and reapplication	Supports the paper's reported results	Together the two sets make assignment changes documented without overstating automation.

Note. This table summarizes the layer distinction; Table S6a quantifies divergence between the rule-hit layer and the released WBS–ITC assignment table.

Table S6a. Per-domain rule-hit-to-assignment diagnostics (N = 443 families × 15 domains = 6,645 binary cells).

Code	Domain name	Rule	Adj.	Added	Removed	Net Delta
1-1	Regolith Processing	47	60	47	34	+13
1-2	Binder / Geopolymer	75	53	13	35	-22
1-3	Composite / Ceramic	65	110	68	23	+45
2-1	Extrusion AM	98	94	14	18	-4
2-2	Powder Bed	22	43	27	6	+21
2-3	Solar / Laser Sintering	16	29	15	2	+13
2-4	Process Monitoring	129	87	21	63	-42
3-1	Autonomous Robots	27	22	11	16	-5
3-2	Teleoperation	9	3	0	6	-6
3-3	Autonomous Construction	24	12	11	23	-12
4-1	Habitat Structures	73	146	76	3	+73
4-2	Shielding	15	27	21	9	+12
4-3	Landing Pads	26	21	5	10	-5
4-4	Deployable Structures	44	65	27	6	+21

Code	Domain name	Rule	Adj.	Added	Removed	Net Delta
4-5	Life Support / ECLSS	7	23	18	2	+16
All	Pooled (6,645 cells)	677	795	374	256	+118

Note. This table reports per-domain transformations from deterministic CPC/title–abstract rule outputs to the released WBS–ITC assignment table over 6,645 patent-family × ITC-domain cells.

Table S7. Candidate extensions to the WBS–ITC specification.

Candidate extension	Evidence in the released materials	Why it sits outside the current specification	Implication for future revision	Indicative priority tier
Excavation	Repeatedly visible in domain-external and near-boundary records	Current specification retains narrow precursor anchors only; it does not provide full excavation coverage.	High priority for future WBS expansion.	High
Foundation engineering	Appears in boundary-set evidence and lower-ranked CPC bridges (e.g., E02D)	Important for translating regolith manipulation into stable structural emplacement.	Would strengthen the transition from materials/process layers to structure deployment.	High
Drilling / boring	Present among transfer-relevant but specification-external cases and CPC bridging signals (e.g., E21B)	Relevant to anchoring, subsurface preparation, and resource access.	Priority extension direction with clear construction relevance.	High
Mixing / emplacement	Observed in boundary-set records associated with in situ preparation steps	Current taxonomy captures fabrication more strongly than emplacement logistics.	Would improve coverage of construction execution tasks.	Medium
Site preparation / surface stabilization	Visible at the edge of landing-pad and regolith-preparation functions	Current specification includes narrow anchors but not a full site-preparation domain.	Important for operational completeness of the construction chain.	High
Road / pad-adjacent surface works	Partly overlaps with 4-3 Landing Pads but not fully exhausted by it	Needed to separate pad construction from broader mobility-surface preparation.	Useful medium-priority refinement once excavation/foundation functions are added.	Medium

Note. Priority map interpreting boundary-set signals under the final versioned specification as structured revision priorities. Tables S7a (CPC-prefix family-count partition) and S7b (pilot separability record) accompany this map; no families are reincorporated into the final 15-domain assignment.

Table S7a. CPC-prefix-based indicative classification of the 70 domain-external families by extension direction. The 70-family count is script-reproducible from the released dataset; the subdivision is an indicative extension-direction partition documented for boundary documentation.

Extension direction	Indicative CPC prefix	Families (n)
Panel A. Specification-external active domain-external families by primary CPC-prefix direction (n = 43, 61.4%)		
Foundation engineering	E02D	28
Crushing / size reduction	B02C	10
Mixing / emplacement	B28C, B01F	3
Size separation / beneficiation	B07B	1
Drilling / boring	E21B	1
Panel B. Boundary-adjacent families (n = 22, 31.4%)		
Cementitious / forming (near ITC 1-2, 1-3)	C04B, B28B	12
Space vehicle / mission design	B64G	6
Formwork / scaffolding (near ITC 4-1)	E04G	4
Unclassified	-	5
Total	-	70

Note. Mutually exclusive partition based on first-matched primary 4-character CPC prefix. Panel A: extension directions outside the final versioned specification (some prefixes such as B02C and B07B serve as narrow anchors elsewhere; these families were not assigned into those domains under the ITC domain criterion). Panel B: directions whose prefixes overlap broader ITC domains but whose families were excluded at assignment. Five families remain unclassified. The CPC-prefix partition does not map one-to-one onto Table S7 priority headings (e.g., E02D spans

foundation, excavation, embankment cases). The total n = 70 active domain-external families is deterministically reproducible from the released dataset (rows with status_category = domain_external_active); the further subdivision into extension-direction categories incorporates a claim-based capability-fit assessment and is provided here as an indicative partition rather than a fully script-generated table. Read with Tables S7 and S7b. A coarse CPC-prefix diagnostic is implemented in scripts/domain_external_partition.py.

Table S7b. Candidate-domain assignment table at the pilot separability review.

Candidate domain	Decision at pilot separability review	Linked retained domain (if any)	Primary reason	Decision note
Dust mitigation	Absorbed	1-1 Regolith Processing	Provisional CPC footprint and first-claim language were not stably separable from regolith-conditioning functions at the selected granularity.	Retained only through narrow precursor anchors inside 1-1.
General environmental control	Absorbed	4-5 Life Support / ECLSS	Claim language and provisional CPC anchors overlapped with habitat-operability support within realized infrastructure systems.	Not retained as a standalone domain in the final versioned specification.
Radiation shielding	Retained	4-2 Shielding	Anchor set and first-claim language remained discriminable during pilot review.	Retained as a standalone domain in the final 15-domain field boundary.
Excavation	Withheld to extension priority	-	Direct construction relevance was recognized, but reproducible CPC/claim separability was not yet stable at the selected granularity.	Remains a high-priority extension direction under the final versioned specification.
Foundation engineering	Withheld to extension priority	-	Strong emplacement relevance, but unstable separation from site-preparation and landing-surface records during pilot review.	Remains a high-priority extension direction.
Drilling / boring	Withheld to extension priority	-	Construction relevance was clear, but pilot signals were limited and boundary-crossing across adjacent construction-chain functions.	Retained as an explicit extension direction rather than absorbed.
Mixing / emplacement	Withheld to extension priority	-	Construction relevance was plausible, but coverage was inconsistent across fabrication and site-work claims at the chosen resolution.	Retained as a medium-priority extension direction.
Site preparation / surface stabilization	Withheld to extension priority	Partial narrow anchoring via 1-1 and 4-3	Partial relevance was visible, but a standalone domain could not be reproducibly delimited during the pilot.	Remains a high-priority extension direction under the final versioned specification.

Note. Decision categories distinguish retained standalone domains, absorbed boundary functions, and withheld extension directions at the pilot separability review. The table documents the final versioned specification and does not alter the full-sample assignment.

Table S7c. 20-family pilot function scope.

Pilot function	Used for	Not used for
20-family pilot	CPC discriminability; first-claim separability; boundary-function stress test (retain/absorb/withhold decisions in Table S7b)	Portfolio share, Jaccard ranking, CPC centrality, domain distribution, top-pair outcomes, jurisdictional composition

Note. The 20-family pilot was purposively sampled across all four WBS layers and was used solely to test domain separability before the specification versioning. Full-sample portfolio, adjacency, and centrality outputs were computed after the specification was versioned; the pilot was used for separability review and not to produce these reported empirical patterns.

Table S8. CPC-anchor sensitivity for top-10 Jaccard co-tagging pairs (N = 373 ITC-classified families).

Rank	Pair	J	CPC anchor tested	N rem.	J adj.	Delta	Interpretation
1	2-1 <-> 2-4	0.341	B33Y (AM class)	39	0.073	-0.268	Strongly conditioned
2	4-1 <-> 4-4	0.294	E04H (structures)	26	0.161	-0.134	Attenuated; residual overlap
3	4-2 <-> 4-4	0.260	B64G (space vehicles)	10	0.143	-0.117	Attenuated; residual overlap
4	2-1 <-> 2-2	0.202	B33Y (AM class)	22	0.011	-0.191	Strongly conditioned
5	1-3 <-> 2-1	0.172	B33Y (AM class)	23	0.046	-0.126	Strongly conditioned
6	1-3 <-> 2-3	0.168	C04B (ceramics)	4	0.139	-0.029	Least attenuated
7	1-1 <-> 2-4	0.167	G01N (testing)	16	0.045	-0.121	Strongly conditioned
8	1-3 <-> 4-1	0.164	E04B (construction)	14	0.107	-0.057	Attenuated; residual overlap
9	4-2 <-> 4-5	0.163	B64G (space vehicles)	6	0.027	-0.136	Strongly conditioned
10	4-1 <-> 4-2	0.161	E04H (structures)	14	0.074	-0.087	Moderately conditioned

Note. Computation: for each domain pair (A, B) and tested CPC prefix c, the adjusted Jaccard is computed after removing from both domain sets all families carrying prefix c—i.e., on the residual family sets F_A^(-c) and F_B^(-c)—as J_c(A,B) = |F_A^(-c) ∩ F_B^(-c)| / |F_A^(-c) ∪ F_B^(-c)|. Notation: J = raw Jaccard; CPC anchor tested = pre-specified theory-guided CPC prefix removed; N rem. = number of intersection families removed; J adj. = post-removal Jaccard; Delta = J adj. - J. Residual values indicate co-tagged domain-pair overlap after removal of one tested anchor.

Table S9. Rule-hit/assignment agreement within the 373-family ITC-classified subset (795 final WBS–ITC tags).

Classification layer	Shared tags with WBS–ITC assignment	Rule-only tags	Rule-positive tags	Assignment-tag agreement	Rule-positive agreement	Assignment-only tags
CPC-rule only	341	76	417	42.9%	81.8%	—
Keyword-rule only	210	128	338	26.4%	62.1%	—
CPC-or-keyword rule union	421	191	612	53.0%	68.8%	374 (47.0%)

Note. Shared tags are patent-family × ITC-domain cells present in both the rule-hit layer and the WBS–ITC assignment. Rule-only tags are rule-positive cells not retained after claim-based capability-fit assessment. Assignment-only tags are WBS–ITC tags with no corresponding rule hit. These quantities are restricted to the 373-family ITC-classified subset and therefore differ from the 6,645-cell accounting in Table S6a.

Denominator reconciliation. Each reported comparison ratio maps to an explicit base. The 443-family active set and the 373-family ITC-classified subset use different rule-positive bases (677 and 612); their 65-cell difference falls on the 70 active domain-external families.

Reported quantity	Count	Base (denominator)	Share
Shared tags (rule hit and assignment agree)	421	795 final assignment tags	53.0%
Assignment-added tags	374	795 final assignment tags	47.0%
Rule-positive cells not retained, active 443-family	256	677 rule-positive cells	37.8%
Rule-positive cells not retained, 373-family subset	191	612 rule-positive cells	31.2%
CPC-only layer agreeing with assignment	341	795 final assignment tags	42.9%
CPC-only tags retained after claim-based capability-fit assessment	341	417 CPC-positive tags	81.8%
Keyword-only layer agreeing with assignment	210	795 final assignment tags	26.4%
Keyword-only tags retained after claim-based capability-fit assessment	210	338 keyword-positive tags	62.1%
Shared tags, 373-family subset	421	612 rule-positive cells	68.8%

Table S10. CPC bridging centrality with seed-anchor flag (N = 443 active families).

Panel A. Full top-10 with seed-anchor flag.

Rank	CPC group	Degree centrality	Betweenness	Stage-1 retrieval seed?	ITC specification anchor?	ITC domain(s)
1	B64G	0.424	0.245	Y	Y	4-1, 4-3, 4-4, 4-5
2	E04H	0.303	0.110	Y	Y	4-1, 4-4
3	G01N	0.291	0.154	Y	Y	2-4
4	E04B	0.267	0.087	Y	Y	4-1, 4-2
5	B33Y	0.255	0.085	Y	Y	2-1, 2-4
6	C04B	0.218	0.096	Y	Y	1-2, 1-3
7	E02D	0.200	0.072	N	N	-
8	H02J	0.188	0.042	N	N	-
9	Y02E	0.182	0.030	N	N	-
10	E21B	0.170	0.029	N	N	-

Panel B. Non-ITC-anchor CPC bridging centrality top-10.

Non-ITC-anchor rank	CPC group	Degree centrality	Betweenness	Stage-1 retrieval seed?	Overall rank
1	E02D	0.200	0.072	N	#7
2	H02J	0.188	0.042	N	#8
3	Y02E	0.182	0.030	N	#9
4	E21B	0.170	0.029	N	#10
5	Y02P	0.158	0.042	N	#13
6	E21C	0.152	0.024	N	#14
7	Y02A	0.152	0.023	N	#15
8	H02S	0.133	0.012	N	#17
9	B32B	0.127	0.015	N	#18
10	F04B	0.115	0.011	N	#19

Note. Degree and betweenness centrality were computed on the binary CPC co-occurrence graph using 4-character CPC prefixes across the 443 active families, with betweenness calculated by Brandes' algorithm. "Stage-1 retrieval seed?" denotes whether the CPC group appeared in the Stage-1 CPC-anchored Lens.org retrieval query. "ITC specification anchor?" denotes whether the same 4-character CPC prefix appears as a domain operationalizer in the final versioned WBS–ITC specification. Non-ITC-anchor codes listed in Panel B are the least directly ITC-anchor-conditioned portion of the bridging profile, but they remain retrieval-route conditioned because the 443-family active set was assembled through CPC-informed retrieval, construction-relevance assessment, and rescue. Reproduced by the CPC bridging script in the released package.

Appendix S1. Interpretation and reconstruction note

Unless otherwise stated, supplementary diagnostics support interpretation of the released WBS–ITC assignment and reconstruction of reported outputs. Co-tagging and bridging values describe domain co-occurrence within the released assignment and active set.

Tables S6a and S9 reconcile denominator-specific rule-hit-to-assignment quantities; Tables S7–S7c document candidate extensions and pilot separability; Tables S11 and S17 summarize reconstruction scope and supported uses.

Table S11. Reconstruction scope of the released package.

Use supported by this study	Supported interpretation	Evidence released	Outside released reconstruction
Computational reconstruction	Reconstruction of reported tables, figures, and sensitivity outputs from the WBS–ITC assignment	Released dataset, scripts, machine-readable specification, query provenance	Live retrieval replay and upstream capability-fit reapplication are documented as provenance
Documented assignment logic	How rule hits differ from WBS–ITC assignments	Rule hits, WBS–ITC assignment table, agreement diagnostics, additions and removals	Alternative field boundaries can be examined in later studies with additional data

Use supported by this study	Supported interpretation	Evidence released	Outside released reconstruction
Later comparison	Released materials can be used to compare alternative retrieval routes, automated assignment methods, or assignment protocols	Dataset, specification, rule hits, WBS–ITC assignment table, scripts	External reference sets and second-domain results can be examined in later studies

Table S12. Descriptive profiles of selected high-overlap intersection sets in the 373-family ITC-classified subset.

Adjacency pair	J	Intersection size	Dominant CPC	Core technology keywords	2020 or later
2-1 <-> 2-4	0.341	46	B33Y, B29C, B22F	3D printing + process monitoring; simulated lunar soil AM; additive manufacturing QA	89%
4-1 <-> 4-4	0.294	48	E04B, B64G, E04H	Inflatable/unfolding habitat + thermal insulation; pressurized cabin systems	83%
4-2 <-> 4-4	0.260	19	B64G, E04B, B32B	Multi-layer inflatable structures; radiation-thermal co-protection shells	68%
2-1 <-> 2-2	0.202	23	B22F, B33Y, B29C	Powder metallurgy + extrusion hybrid; laser sintering of lunar soil composites	91%
1-3 <-> 2-1	0.172	30	B33Y, C04B, E04B	Ceramic/geopolymer feedstock for 3D printing; regolith-based construction materials	90%
1-3 <-> 2-3	0.168	20	C04B, B28B, G01N	Microwave/solar sintering of lunar soil bricks; energy-based ceramic consolidation	100%

Note. Each row summarizes the patents co-tagged to both ITC domains, their dominant CPC prefixes, core technology keywords extracted from titles and abstracts, and a qualitative recency descriptor. "2020 or later" denotes the share of intersection-set families whose earliest priority year is 2020 or later. These descriptions are co-tagging summaries within the released WBS–ITC assignment (claim boundaries: Table S17).

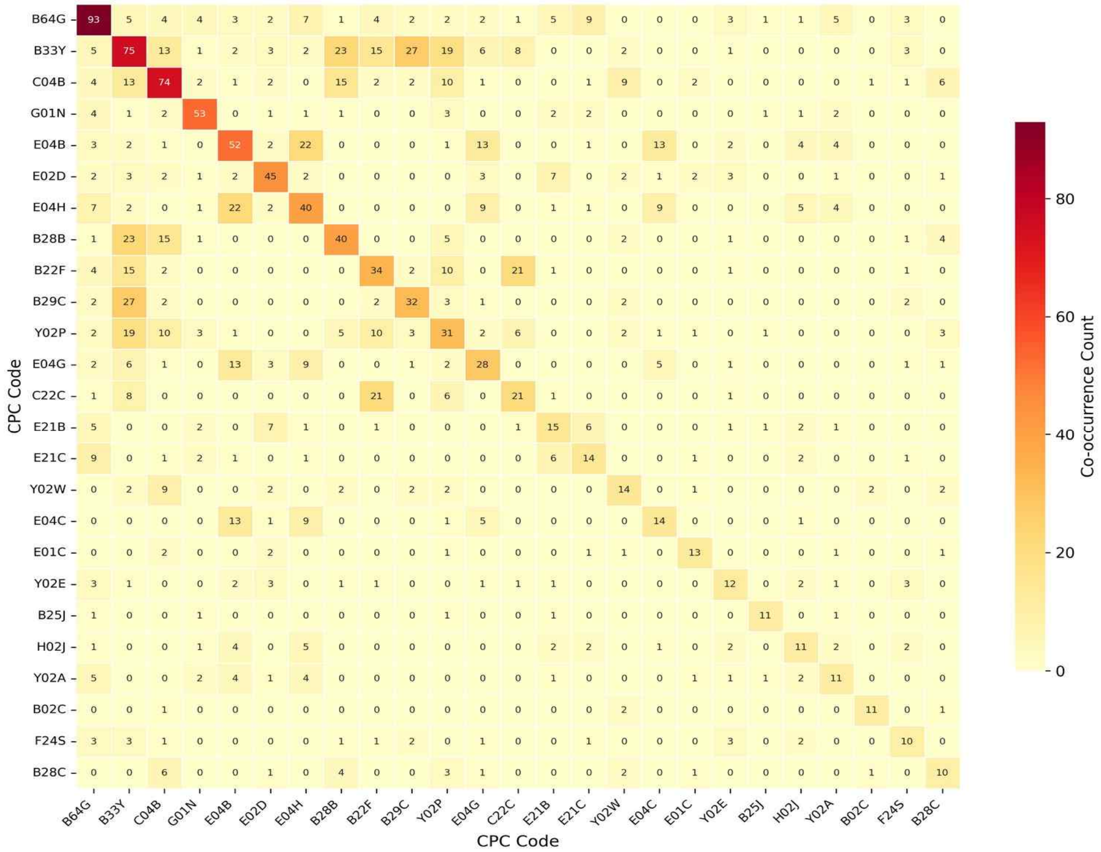

Fig. S1. CPC co-classification heatmap for the 25 most prevalent four-character CPC prefixes. Heatmap color intensity represents the number of patent families sharing both CPC group prefixes;

the top 25 most prevalent prefixes in the 443-family active set are shown. Generated from the CPC co-classification matrix for the 443-family active set; the corresponding bridging-centrality results are reported in Table S10.

Table S13. Rescue-route sensitivity check for portfolio, Jaccard, and CPC-bridging outputs.

Diagnostic	With rescue-route classified families	Excluding nine classified rescue-route families	Interpretation
Active set	443 families	434 families	The documentation-only below-threshold rescue record is excluded from computations in both cases.
ITC-classified subset	373 families	364 families	Same final versioned WBS–ITC specification; no domain, rule, or boundary definition changes.
WBS-layer tag shares	Baseline reported in the manuscript	No WBS-layer share changes by more than 0.5 percentage points	Rescue-route families do not drive the portfolio-level WBS distribution.
Top six Jaccard pairings	Baseline top-six pairings reported in Table 8	Same top six pairings preserved	Rescue-route families do not drive the leading specification-internal co-tagging order.
CPC-bridging leading trio	B64G, E04H, G01N	Same leading trio preserved, with only minor rank-order variation	Rescue-route families do not drive the main CPC-bridging profile.

Note. This table reports a route-sensitivity check under the final versioned specification. It evaluates whether the supplementary CPC-prefix retrieval pass changes the main downstream patterns. The rescue pass introduced no new WBS layers, ITC domains, CPC anchors, keyword rules, or assignment criteria.

Table S14. Within-layer comparison of deterministic first-pass and final capability-assignment mapping outputs on the 443-family active set (see manuscript Fig. 5 and Table 9).

Panel A. Portfolio coverage and WBS-layer tag-share distribution.

Delineation layer	N classified	Total tags	WBS-1 %	WBS-2 %	WBS-3 %	WBS-4 %
CPC-rule only	332	452	23.7	39.8	9.7	26.8
CPC-or-keyword (rule union)	393	677	27.6	39.1	8.9	24.4
Final assignment (record)	373	795	28.1	31.8	4.7	35.5

WBS shares are computed as (WBS-layer tag count) / (total tags) within each delineation layer. CPC-rule and rule-union counts are produced by deterministic matching against itc_codebook.json on the same 443-family active set; the WBS–ITC assignment is the WBS–ITC assignment table (the main-text Results section).

Panel B. Top-6 Jaccard co-tagging pairs by delineation layer.

Delineation layer	Pair	Jaccard J	Intersection size n
CPC-rule only	3-3 <-> 4-1	0.162	12
CPC-rule only	2-1 <-> 2-2	0.136	12
CPC-rule only	4-1 <-> 4-4	0.127	10
CPC-rule only	2-1 <-> 2-4	0.124	17
CPC-rule only	4-1 <-> 4-2	0.111	7
CPC-rule only	1-3 <-> 2-1	0.106	13
CPC-or-keyword (rule union)	4-1 <-> 4-4	0.245	23
CPC-or-keyword (rule union)	2-1 <-> 2-2	0.154	16
CPC-or-keyword (rule union)	3-3 <-> 4-1	0.141	12
CPC-or-keyword (rule union)	1-2 <-> 2-1	0.138	21
CPC-or-keyword (rule union)	2-1 <-> 2-4	0.129	26
CPC-or-keyword (rule union)	1-3 <-> 2-1	0.124	18
Final assignment (released)	2-1 <-> 2-4	0.341	46
Final assignment (released)	4-1 <-> 4-4	0.294	48
Final assignment (released)	4-2 <-> 4-4	0.260	19
Final assignment (released)	2-1 <-> 2-2	0.202	23

Delineation layer	Pair	Jaccard J	Intersection size n
Final assignment (released)	1-3 <-> 2-1	0.172	30
Final assignment (released)	1-3 <-> 2-3	0.168	20

Interpretation. Relative to deterministic rule-only delineation, the released WBS–ITC assignment shows a materially different WBS profile: Manufacturing share contracts (39.8% -> 31.8%), Robotics contracts further (9.7% -> 4.7%), and Structures expands (26.8% -> 35.5%). The leading manufacturing-monitoring Jaccard value (2-1 <-> 2-4) is 2.75x larger in the WBS–ITC assignment (J = 0.124 -> 0.341), and the habitat-envelope Jaccard value (4-1 <-> 4-4) is 2.31x larger in the WBS–ITC assignment (J = 0.127 -> 0.294). The CPC-rule-only top pair 3-3 <-> 4-1 (J = 0.162) does not appear in the final-assignment top-6, illustrating capability-fit removal of broad autonomous-construction rule hits when the underlying claim describes manual or passive emplacement rather than autonomous robotic execution.

Note. This table reports a within-layer delineation-strategy comparison on the same 443-family analytically active set. The "Final assignment" column is the WBS–ITC assignment for this study, as bounded in Table S17 and the Discussion section of the main text; this comparison illustrates what claim-based capability fit adds and removes relative to the deterministic first-pass. Reproduced by scripts/seed_only_baseline.py.

S15. Comparison materials note

The released materials define comparison materials consisting of family records, CPC codes, the WBS–ITC specification, deterministic rule outputs, the WBS–ITC assignment table, and reconstruction scripts. Later studies may compare alternative delineation, retrieval, assignment protocols, or external reference sets using these materials.

Table S16. Expanded within-corpus comparison against deterministic and restrictive assignment variants.

This table extends the within-layer comparison reported in Table S14 by adding a keyword-rule-only baseline and an explicit-lunar/regolith subset (rule-union restricted to families whose title or abstract contains explicit lunar or regolith terms), in addition to the CPC-rule-only, rule-union, and WBS–ITC assignment variants. The five variants are bounded operational delineations on the same 443-family active set; the final WBS–ITC assignment is represented by the WBS–ITC assignment table. The numeric values below are regenerated by scripts/expanded_baseline.py.

Layer	N cl.	Tags	Assignment-tag overlap	Variant-positive overlap	Top-6 ov.
CPC-rule only	332	452	42.9%	75.4%	4/6
Keyword-rule only	271	377	26.4%	55.7%	0/6
CPC-or-keyword (rule union)	393	677	53.0%	62.2%	4/6
Explicit-lunar subset (rule-union restricted)	249	445	39.4%	70.3%	3/6
Final assignment (released)	373	795	Same layer	Same layer	6/6 (reference)

Note. "N cl." is the number of classified active families under each variant. "Assignment-tag overlap" is the share of WBS–ITC assignment tags also present in the variant layer. "Variant-positive overlap" is the share of variant-positive tags also present in the WBS–ITC assignment. "Top-6 ov." is the number of the WBS–ITC assignment top-6 Jaccard pairs that also appear in the variant top-6. The WBS–ITC assignment row is the same layer compared with itself. Overlap values in this table are computed over variant-positive tags on the 443-family active set and therefore are not numerically identical to the S9 values, which are restricted to the 373-family ITC-classified subset.

Appendix S2. Lens.org retrieval query anatomy

The initial retrieval universe was constructed through a two-stage CPC-based and keyword-expanded query on Lens.org (retrieval date: 2026-03-28; export format: JSON records with Lens simple-family identifiers retained at the record level, with all downstream analyses consolidated to the simple-family level). The full provenance file (lens_query.txt) is included in the reproduction package.

Query component 1 — CPC-anchored retrieval. CPC prefixes used in primary search: B64G (cosmonautic vehicles/equipment), C04B (lime; magnesia; slag; cements; ceramics; refractories), B33Y (additive manufacturing), E04H (buildings for particular purposes; shelters), B22F (working metallic powder), B28B (shaping clay or other ceramic compositions), G01N (investigating/analysing materials), B25J (manipulators; robots), E04B (general building constructions).

Query component 2 — Keyword expansion. Boolean structure: ("lunar" OR "regolith" OR "planetary" OR "extraterrestrial" OR "moon" OR "mars" OR "space construction" OR "in-situ resource") AND ("construction" OR "manufacturing" OR "additive" OR "structure" OR "material" OR "sintering" OR "habitat" OR "shielding" OR "3d print" OR "robotic" OR "landing pad"). Additional targeted queries: "contour crafting" OR "lunar concrete" OR "regolith processing" OR "extraterrestrial construction" OR "space manufacturing".

Note on robotics sensitivity. A robotics-expanded query variant, adding broader B25J subclasses and autonomous-construction keywords, was tested during exploratory retrieval-sensitivity checks before the final retrieval query. It returned only a small number of marginal candidates after de-duplication, nearly all of which were already captured through co-classification with B64G or E04H seeds. Because the incremental yield did not justify the additional retrieval-route complexity, the expanded variant was not adopted for the final retrieval. The narrower claim-supported representation of the Robotics layer (WBS-3; ITC 3-1, 3-2, and 3-3; 4.7% of total tags) therefore reflects the operational scope of the 2026-03-28 retrieval query rather than an ontological exclusion of robotic construction from the ISRU field.

Post-retrieval processing. Deduplication at the simple-family level; two-pass construction-relevance assessment using predefined criteria; supplementary retrieval ("rescue") pass using supplementary CPC-prefix cues to identify additional families (10 rescue families, 9 ITC-classified). Construction-relevance assessment determined whether records entered the active set. Domain tags were assigned later using the representative abstract and first independent claim. Note: The analytical dataset is not a direct Boolean output. The final 453-family released dataset was constructed through staged construction-relevance assessment, WBS–ITC assignment, and supplementary retrieval. Exact retrieval-universe replay requires the Lens.org saved search state at the retrieval date, which may not be reproducible due to index updates and coverage changes.

Retrieval flow record counts for the construction-core corpus.

Step	Operation	Count
R1	Initial Lens.org retrieval (CPC-anchored prefixes + keyword expansion)	1,093 records
R2	Two-pass construction-relevance assessment using predefined criteria (retention threshold >= 2.5; no AI-generated ranking was used to exclude records)	511 records
R3	Supplementary ITC-oriented rescue from records not retained after the first construction-relevance pass	+12 records
R4	Pre-deduplication corpus	523 records
R5	Lens Simple Patent Family deduplication (37 multi-record groups; 70 jurisdictional duplicates)	453 unique families
R6	Status-flag exclusions (9 false matches; 1 documentation-only below-threshold record)	-10 families
R7	Analytically active set	443 families
R8	ITC-classified subset (>= 1 final-assignment tag)	373 families
R9	Final WBS–ITC tags	795 tags

Note. Retrieval-flow step R3 is reported at the pre-deduplication level; of these 12 rescued records, 2 were absorbed into existing simple patent families during step R5, so 10 independent rescue families appear in the final 453-family released corpus. The 80 unclassified families in the 453-family corpus comprise 70 active domain-external families with plausible construction relevance but no adequate fit within the final 15-domain ITC specification, 9 false

matches with no plausible construction-relevance pathway, and 1 documentation-only below-threshold record retained only for construction-relevance documentation. The step R1 record count is from the 2026-03-28 retrieval; exact replay of the Lens.org saved search state is not claimed because index coverage may change over time.

Construction-relevance criteria. The anchors below document construction-relevance scores used to decide whether records entered the active set; they are not WBS–ITC assignment categories.

Construction-relevance score	Construction-relevance meaning
0	No plausible construction-chain transfer relevance
1	Remote analogy only; no claim-supported construction relevance
2	Possible but weak construction relevance; normally below retention threshold
3	Plausible construction relevance to ISRU construction
4	Strong construction-chain relevance adjacent to WBS–ITC scope
5	Direct claim-supported construction-chain relevance within or immediately adjacent to WBS–ITC scope

Chronology note. The WBS–ITC classification study and patent-search design began in February 2026. The primary Lens.org retrieval was executed on 2026-03-28, and the retrieval query provenance is released as data/lens_query.txt. The WBS–ITC specification was developed during the study and is released as data/itc_codebook.json (version 2.0). The released file records the domain definitions, CPC anchors, and keyword stems used for the released WBS–ITC assignment table; its checksum in MANIFEST.md lets readers confirm that the released specification is the one used.

Table S17. Uses supported by this WBS–ITC procedure and additional analysis needed for broader claims.

Object	Use supported by this study	Additional analysis needed for broader claims
A1. Study type	A patent-based capability map of lunar ISRU space construction, with a WBS–ITC domain-assignment basis that can be reused as a reporting structure for classification-dispersed patent fields.	Landscape, forecast, maturity-ranking, market-opportunity, and domain-census readings can be examined with broader or longitudinal designs.
B1. Capability architecture	A bounded WBS–ITC capability architecture whose domains are derived from national and space-agency space-construction roadmaps and released in the final versioned specification.	Complete lunar-construction or all-ISRU ontologies require a broader specification.
B2. CPC conditioning	CPC anchors are visible candidate-generation devices whose conditioning is reported.	CPC conditioning is documented as part of the candidate-generation layer; it does not define the final WBS–ITC assignment.
B3. Rule-hit layer	CPC/title–abstract rules provide transparent and computationally reproducible candidate generation.	Rules are used as transparent candidate generation.
B4. WBS–ITC assignment	The WBS–ITC specification is applied to representative abstracts and first independent claims to produce the assignments used for reported results.	Alternative boundaries can be compared using the released dataset, rule hits, assignment table, and scripts.
B5. WBS–ITC assignment table	Retained, removed, added, and jointly absent patent-family × ITC-domain cells document how rule hits are converted into WBS–ITC tags.	Alternative assignment rules can be examined in later studies with additional data.
C1. Agreement/divergence diagnostics	Rule-hit/assignment agreement, assignment-only tags, and rule-positive hits not retained.	Reports rule-hit-to-assignment differences used in the present comparison.
C2. Agreement/divergence magnitudes (53.0% agreement; 374 assignment-only / 256 rule-positive not retained; +118 net)	Case-specific procedure outputs to be re-measured per domain.	Gap magnitudes are case-specific quantities to be re-measured per field.
C3. Adjacency outputs (Jaccard, CPC bridging)	Co-presence and co-tagging indicators computed within the WBS–ITC assignment.	Temporal, engineering, causal, and maturity claims require additional data beyond patent co-tagging.

Object	Use supported by this study	Additional analysis needed for broader claims
C4. Portfolio & jurisdiction	A representational pattern within a construction-scoped, CN-dominant (~75%) active set.	Jurisdiction-balanced census and field-maturity assessment can be studied with different designs.
C5. Domain-external set (70 families)	Boundary evidence and candidate extension directions under the WBS–ITC specification.	Domain-external families remain outside the current 15-domain specification.
C6. Component and nested boundary-variant analysis	Component and nested deterministic variants and the final-assignment transformation are summarized on the same admitted corpus to illustrate boundary sensitivity.	Provides within-corpus comparison across operational variants.
D1. Released materials	Computational reconstruction of reported outputs from the dataset, specification, rule hits, WBS–ITC assignment table, and scripts.	Live retrieval replay and reapplication of the upstream capability-fit step require additional provenance or additional data.
D2. Transferability	Later studies can repeat the same sequence after rebuilding the domain set for another classification-dispersed field.	The procedure transfers after domain-specific capability architecture, stop rules, candidate domains, and assignment criteria are rebuilt.

Supplementary References

Cesaretti, G., Dini, E., De Kestelier, X., Colla, V., & Pambaguian, L. (2014). Building components for an outpost on the Lunar soil by means of a novel 3D printing technology. *Acta Astronautica, 93*, 430-450. [Cross-referenced to main reference list]

Heemskerk, M. V., Van Westrenen, W., & Foing, B. H. (2020, March). Lunar and Martian regolith based concrete as building blocks for future human settlements. In *51st Annual Lunar and Planetary Science Conference* (No. 2326, p. 2878).

KICT. (2023). Space architecture: Development of core technology for the construction of lunar habitation (1/3). KICT 2023-050. [Cross-referenced to main reference list]

KICT. (2024). Space architecture: Development of core technology for the construction of lunar habitation (2/3). KICT 2024-061. [Cross-referenced to main reference list]

Khoshnevis, B., Bodiford, M. P., Burks, K. H., Ethridge, E., Tucker, D., Kim, W., Toutanji, H., & Fiske, M. R. (2005). Lunar contour crafting: A novel technique for ISRU-based habitat development. *43rd AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada. AIAA 2005-538. https://doi.org/10.2514/6.2005-538

Lim, S., Prabhu, V. L., Anand, M., & Taylor, L. A. (2017). Extra-terrestrial construction processes: Advancements, opportunities, and challenges. *Advances in Space Research, 60*(7), 1413-1429. [Cross-referenced to main reference list]

Meurisse, A., Makaya, A., Willsch, C., & Sperl, M. (2018). Solar 3D printing of lunar regolith. *Acta Astronautica, 152*, 800-810. [Cross-referenced to main reference list]

Miaja, P. F., Navarro-Medina, F., Aller, D. G., Leon, G., Camanzo, A., Suarez, C. M., Alonso, F. G., Nodar, D., Sauro, F., Bandecchi, M., et al. (2022). RoboCrane: A system for providing a power and a communication link between lunar surface and lunar caves for exploring robots. *Acta Astronautica, 192*, 30-46. https://doi.org/10.1016/j.actaastro.2021.11.023

Mueller, R. P. (2022). Lunar base construction overview. NASA Kennedy Space Center, Swamp Works, Exploration and Research Technologies Division. [Cross-referenced to main reference list]

Mueller, R. P., Howe, S., Kochmann, D., Ali, H., Andersen, C., Burgoyne, H., Chambers, W., Clinton, R., De Kestellier, X., Ebrahimi, K., Fateri, M., Howe, A., Leach, N., Martins, P., Nakamura, T., & Werkheiser, N. (2016). Automated additive construction (AAC) for Earth and space. In *Proceedings of the Fifteenth Biennial ASCE Conference*. https://doi.org/10.1061/9780784479971.036 [Cross-referenced to main reference list]

Okonkwo, C., Awolusi, I., Bin-Shafique, S., Ximenes, S., Gorin, B., Redfield, J., & Petersen, C. (2026). Hybrid simulation framework for autonomous robotics lunar launch and landing pad construction. [Cross-referenced to main reference list]

Reiss, P., Grill, L., & Barber, S. J. (2019). Sintering of lunar regolith simulant under vacuum conditions. In *Earth and Space 2019*. ASCE. [Cross-referenced to main reference list]

Sanders, G. B., Kleinhenz, J. E., & Hilburger, M. W. (2024). Using ISRU and surface construction to define long-term lunar infrastructure needs. *AIAA AVIATION Forum and ASCEND 2024*, AIAA 2024-4840. https://doi.org/10.2514/6.2024-4840 [Cross-referenced to main reference list]