RISC-V Unified Database [1]

Afonso Oliveira[1]

[1]Synopsys, Inc.

Abstract

We present the RISC-V Unified Database (UDB), a single, machine-readable source-of-truth that consolidates the RISC-V specification into a unified framework. As the RISC-V ecosystem expands—with nearly 200 ratified extensions—the specification has become increasingly fragmented and challenging to verify, particularly for safety-critical domains such as space exploration. UDB addresses these challenges by aggregating all technical details into a rigorously validated database, enabling the automated generation of essential artifacts such as ISA manuals, instruction and CSR indices, stringent certification documents and even an Instruction Set Simulator (ISS). This unified approach streamlines tool creation and data verification, thereby empowering the development and certification of space-qualified processors. Although significant progress has been achieved, UDB remains a work in progress, and we invite active community participation to further enhance its capabilities.

Introduction

Space systems require computing platforms that can operate reliably in extreme environments. From satellite control to deep-space exploration, processors must withstand radiation exposure, temperature fluctuations, and power constraints while maintaining high performance and fault tolerance [2, 3]. Traditional space computing has long relied on proprietary architectures with high costs and long development cycles, limiting accessibility and innovation. As the computational demands of space missions continue to grow—driven by advancements in artificial intelligence, autonomous navigation, and high-speed data processing—there is a pressing need for flexible and sustainable processor architectures.

RISC-V has become a compelling choice for space applications due to its open, modular, and customizable architecture. Its flexibility allows different space applications to use tailored processors optimized for specific workloads, rather than relying on one-size-fits-all designs. Additionally, its open-source nature is crucial for the collaborative nature of space exploration, where government agencies, research institutions, and private aerospace companies from different countries must work together. With organizations like NASA, ESA, and major industry players actively exploring or adopting RISC-V [4], it provides a shared, vendor-neutral platform that speeds up development, improves interoperability, and ensures long-term sustainability.

However, as RISC-V adoption accelerates, its rapidly growing ecosystem presents new challenges in managing and utilizing its vast and evolving body of information. With nearly 200 ratified extensions[5], the specification is distributed across multiple repositories and formats, making it difficult to efficiently track architectural components, extensions, and interdependencies. The intricate relationships between instructions, Control and Status Registers (CSRs), and extensions add to the complexity, making it increasingly challenging to extract accurate and up-to-date insights—particularly in safety-critical applications such as space, where precision is paramount.

To address these challenges, the RISC-V Unified Database (UDB) serves as a centralized, structured source of truth for the RISC-V specification. It consolidates critical technical details—including the full ISA specification, instruction and CSR descriptions, and architectural profiles—into a machine-readable format. This structured approach enables rapid data retrieval, automated validation, and seamless cross-referencing, ensuring accuracy and consistency across development workflows. By providing a reliable foundation for generating tailored, verified outputs, UDB streamlines the creation of certification and verification tools, as well as customized documents such as profile manuals and certification requirements. Leading industry players, including Qualcomm and Synopsys, along with the RISC-V Certification Steering Committee (CSC), have already integrated UDB into their workflows, leveraging its efficiency and precision to enhance development processes.

For space applications, UDB offers a powerful framework to simplify the development, verification, and qualification of RISC-V-based processors. By centralizing architectural insights and ensuring consistency across information, it allows engineers and researchers to make informed, data-driven decisions while maintaining the rigorous standards required for mission-critical computing. Additionally, UDB serves as a foundation for building space exploration tools, serving as the base-layer for the creation of specialized software for validation, simulation, and certification. As RISC-V continues to grow, UDB is positioned to become a key enabler for the next generation of high-performance, space-qualified processors and tools.

Background

Challenges with RISC-V Specification Ecosystem

The RISC-V specification ecosystem currently relies on multiple disconnected repositories and specification formats. Key documents, such as the ISA and assembly manuals, are written in AsciiDoc – a format not meant for structured data capture – forcing reliance on copy-pasting or rewriting for derivative document generation and precluding data verification.

The intricate relationships between instructions, extensions, Control and Status Registers (CSRs), architectural parameters (most of which are unnamed), and profiles further amplify the complexity of these challenges. Instructions belong to one (or more) extensions, yet tracking which instructions are introduced, modified, or forbidden across different extensions is difficult. CSR interactions with specific extensions and privilege levels are often unclear. Profiles group extensions into structured subsets of the ISA, but the lack of consistent cross-referencing makes it difficult to determine which components apply to a given subset.

Version management further compounds this complexity. A typical CPU design incorporates multiple extensions, each evolving independently as new versions are ratified. For example, a current design may specify `RV32IM_Zba1p0_Zbb1p0_Zbs1p0_Zicsr2p0`, but as the ecosystem progresses beyond "1.0" and "2.0" versions, maintaining consistency across implementations and documentation will become more difficult.

All of this creates challenges for the entire ecosystem. For example, it's difficult for vendor documentation to stay up-to-date with the standard given the reliance on copy/paste. Tool developers (*e.g.*, compilers) must create and maintain their own machine-readable representations. And implementations struggle to verify a core against documentation that is not easily understood or referenced.

RISC-V Unified Database

UDB organizes information systematically into distinct categories, such as extensions, instructions, CSRs, architectural parameters, profiles, and certificates, each characterized by precise attributes and clearly defined relationships. Additionally, UDB facilitates customization through user-defined overlays that extend or modify the base specification without altering its core data.

UDB data is represented using YAML, conforming strictly to a schema enforced by JSON Schema. YAML is chosen for its readability by both humans and machines, as well as its broad support across mainstream programming languages. Each UDB component is modularly structured, incorporating explicit attributes that define its properties and interconnections. This modular design allows entities to be referenced, inherited, or extended seamlessly without redundant duplication. For instance, instructions explicitly *refer* to their defining extensions, profiles *include* ISA extensions, and architectural parameters delineate specific implementation choices. The schema also inherently supports features like versioning, dependency resolution, and formal definitions (such as IDL and Sail), making it highly suitable for generating structured documentation, configuring Instruction Set Simulators (ISSs), and enabling integration with various toolchains.

Artifact	Current	UDB
ISA Manuals	Prose	Prose + Appendices
ISS / Formal model	Incomplete	Generator complete, incomplete semantics
Opcode Database	Incomplete	Complete (superset of `riscv-opcodes`)
Profile Manuals	Prose	Prose + Appendices
CSC Documents	Using UDB	-
Instruction Index	Non-existent	Complete mnemonics/encoding, Incomplete semantics/descriptions
CSR Index	Non-existent	Incomplete

Table 1: *Comparison of current RISC-V artifacts versus UDB*

Table 1 contrasts UDB with existing approaches. Although resources such as profile manuals and opcode databases existed previously, they were either incomplete or manually maintained. Comprehensive instruction and CSR indices, which offer structured cross-referencing capabilities, were mostly absent in existing documentation but are now actively being developed within UDB. By consolidating these various outputs, UDB significantly enhances clarity, accessibility, and long-term maintainability throughout the RISC-V ecosystem.

Data Status. Currently, UDB encompasses encoding data for all ratified RISC-V instructions documented in `riscv-opcodes`, as well as the ratified instruction data utilized by the *GNU* and *LLVM* toolchains. While formal execution semantics have been defined for numerous extensions, some notable exceptions remain, such as the Vector (V) and Hypervisor (H) extensions, which are either incomplete, missing, or pending verification. The UDB community is actively addressing these gaps, backed by identified funding support.

Data Verification. UDB rigorously cross-validates its data against established de-facto reference sources. Encoding information is systematically verified against `riscv-opcodes`, supported by a dedicated UDB generator producing outputs consistent with those of `riscv-opcodes`. This robust validation positions UDB to potentially replace `riscv-opcodes` as the authoritative reference tool, contingent upon community adoption. Moreover, UDB's instruction data undergoes meticulous validation against LLVM's instruction tables, leveraging the comprehensive datasets from LLVM's table generator. This rigorous cross-validation process swiftly identifies and rectifies discrepancies, ensuring that only thoroughly verified data is incorporated into UDB. Future enhancements, including integration of LLVM-compatible disassembly information and relocation data, will further reinforce UDB's standing as a reliable source-of-truth.

UDB Empowering Space Exploration

In the harsh and unforgiving environment of space, ensuring the reliability of every component is critical. The current approach to managing the RISC-V specification may pose challenges and delays for space applications, where even minor discrepancies can lead to mission-critical failures.

UDB revolutionizes this landscape by centralizing the entire RISC-V specification into a single, machine-readable database. This consolidation not only guarantees that the specification data is accurate - due to its verification process - but also streamlines the creation of essential tools. For instance, the UDB-integrated ISS is directly generated from validated data, ensuring that simulations precisely reflect the intended behavior of space-qualified processors and serving as a proof-of-concept for how UDB simplifies tool creation.

By automating the verification and generation of critical artifacts, UDB minimizes manual interventions and reduces the risk of human error—a significant advantage for space systems where reliability is paramount. Moreover, UDB's rigorous data validation process facilitates the rapid production of strict certification documents. These documents are indispensable for meeting the stringent requirements of space missions, where every element must be meticulously verified before deployment.

In essence, UDB empowers space exploration by:

- **Streamlining Tool Creation:** Enabling faster generation of robust tools by serving all the data in a variety of formats, ensuring that simulation and validation are based on the definitive specification.
- **Ensuring Data Accuracy:** Maintaining a single source-of-truth for the RISC-V specification and cross-verifying it with external sources, which guarantees that all derived outputs are consistent and correct.
- **Facilitating Certification:** Allowing for the efficient production of rigorous certification documents and verification tools that are critical for ensuring mission reliability in extreme environments.

By bridging the gap between raw specification data and practical, certified outputs, UDB sets a new standard for developing, verifying, and certifying RISC-V-based processors for space applications. This transformative approach not only reduces development time and cost but also significantly enhances the overall reliability and safety of space-bound computing systems.

Discussion and Call to Action

UDB has garnered initial endorsement from the RISC-V Technical Steering Committee (TSC) and has officially received a dedicated Special Interest Group (SIG) under the RISC-V umbrella. This formal recognition underscores the community's commitment and highlights the significance of UDB within the broader RISC-V ecosystem, further enabling systematic improvement of existing repositories.

UDB is currently collaborating closely with several key RISC-V working groups, including:

- **Documentation SIG** to substantially enhance the quality, coherence, and navigability of RISC-V documentation;
- **CSC** to facilitate the generation of stringent and reliable certification documents essential for safety-critical applications;
- **TSC** to ensure alignment of efforts with overarching community objectives, standards, and practices.

Additionally, UDB engages with key ecosystem contributors such as toolchain maintainers, ISS developers, and hardware design and verification engineers. This inclusive collaboration ensures comprehensive coverage of ISA-related information and accelerates integration of new and critical data. Furthermore, the scope of UDB is expanding to encompass non-ISA data, enhancing its applicability to a broader range of design and verification workflows.

Moving forward, the UDB community is actively focused on several strategic priorities:

- **Completing Data Gaps:** Filling gaps in textual descriptions, formal semantics, and validation processes, while continually enhancing the robustness of UDB infrastructure.
- **Auditable Documentation Trails:** Establishing direct linkage between structured UDB data

and normative text within official prose manuals, creating transparent and auditable reference trails.
- **Extending the Framework:** Adapting and extending the UDB framework to meet emerging demands which require rigorous certification and high-fidelity verification processes.
- **Toolchain Integration Enhancements:** Enhancing data sets to include richer information, such as relocation data, ensuring compatibility and ease of integration with widely-used toolchains and verification tools.

We extend an open invitation to researchers, engineers, space industry stakeholders, and the broader RISC-V community to actively participate in and contribute to the evolution of UDB. Your involvement and expertise will be instrumental in establishing UDB as the definitive, centralized source-of-truth for the RISC-V specification, driving reliability, interoperability, and innovation within mission-critical and safety-critical systems.

This BSD-3 licensed project thrives on community participation, and your contributions will help shape the future of RISC-V, ensuring its continued success and widespread adoption across diverse and demanding application domains.

Conclusion

UDB has been successfully bootstrapped, laying a strong foundation for consolidating the RISC-V specification into a unified, structured, and machine-readable format. By centralizing specification management, UDB significantly enhances consistency, reduces errors, and accelerates the generation of critical artifacts, including ISA manuals, instruction and CSR indices, and simulation models. Its structured, verifiable approach enables robust validation processes, streamlines certification workflows, and provides essential support for the rigorous demands of safety-critical and mission-critical domains. As the RISC-V ecosystem continues to grow and diversify, UDB stands poised to become an essential resource, empowering innovation, reliability, and collaboration across the entire community.

References

[1] Hower, Derek, et al. *The RISC-V Unified Database*. URL: https://github.com/riscv-software-src/riscv-unified-db.

[2] Microchip Technology Inc. "NASA JPL Selects Microchip for Game-Changing Spaceflight Computing Processor". In: (Sept. 2022). URL: https://www.microchip.com/en-us/about/newsroom/nasa-jpl-selects-microchip-for-game-changing-spaceflight-computing-processor.

[3] Alice Chung and Ravi Kumar. "Validation of RISC-V Implementations for Space Applications". In: *Proceedings of the 2023 IEEE Aerospace Conference*. Focuses on validation methods for RISC-V based space systems. IEEE, 2023, pp. 200–207.

[4] Tobias Mann. "SiFive RISC-V CPU cores to power NASA's next spaceflight computer". In: *The Register* (Sept. 2022). URL: https://www.theregister.com/2022/09/06/nasas_spaceflight_computer_risc_v/.

[5] RISC-V International. *The RISC-V Instruction Set Manual, Volume I: Unprivileged ISA*. Version 20240411. 2024.