Artificial Intelligence in Medicine 92 (2018) 7–9

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Historical Perspectives

Origins of the Arden Syntax

George Hripcsak[a,b,*], Ove B. Wigertz[c], Paul D. Clayton[a,d]

[a] Department of Biomedical Informatics, Columbia University, 622 W 168th Street, PH20, New York, NY 10032, USA
[b] Medical Informatics Services, NewYork-Presbyterian Hospital, 622 W 168th Street, PH20, New York, NY 10032, USA
[c] Department of Medical Informatics, Linköping University, 58183 Linköping, Sweden
[d] Department of Biomedical Informatics, University of Utah, 421 Wakara Way, Ste 140, Salt Lake City, UT 84108, USA

ARTICLE INFO

Keywords:
Arden Syntax
Knowledge-based system
Automated decision support

ABSTRACT

The Arden Syntax originated in the 1980's, when several knowledge-based systems began to show promise, but researchers recognized the burden of recreating these systems at every institution. Derived initially from Health Evaluation through Logical Processing (HELP) and the Regenstrief Medical Record System (RMRS), the Arden Syntax defines medical logic that can be encoded as independent rules, such as reminders and alerts, with the hope of creating a public library of rules. It was first vetted at an informatics retreat held in 1989 at Columbia University's Arden Homestead. The syntax was intended to be readable by clinician experts but to provide powerful array processing, which was derived largely a programming language called APL. The syntax was improved and implemented by a number of researchers and vendors in the early 1990's and was initially adopted by the consensus standards organization, the American Society for Testing and Materials.

The Arden Syntax traces its roots to the mid-1980s, when a number of knowledge-based systems had shown promise in the institutions of their inventors, and informaticians wondered how we would ever deploy these systems broadly around the world. One such system was Health Evaluation through Logical Processing (HELP) [1] at LDS Hospital in Salt Lake City. The HELP system was initially commercialized by the Control Data Corporation (CDC; 3M Corporation commercialized it later), making the execution environment sharable but leaving open how to share the knowledge base of rules. In 1985, Paul Clayton, then at LDS Hospital, was invited by Joachim Dudeck of Giessen, Germany, to spend 6 months in Giessen in an endeavor to install the CDC version in their clinic. The university in Giessen had a tandem computer and he received tapes from CDC and a stipend from CDC and the German Research Foundation to take his family of eight children for a 6-month sabbatical. He was able to get some parts of the system working—mainly able to evaluate a few rules—but not much in the way of data collection or data entry [2,3]. In the fall of 1985, he attended the annual conference on "Computers and Cardiology" hosted by Ove Wigertz of Linköping, Sweden. The social program was an orienteering event, and he was paired with Allan Pryor of LDS Hospital, whom he had not seen for 6 months. As they walked through the beautiful woods of Sweden, they discussed the difficulty of moving a customized clinical information system to a new setting. They were especially sensitive to the 15-year effort led by Homer Warner of LDS Hospital that had by that time gone into developing, evaluating, and refining the rules used in the HELP system. They felt that the HELP system and a similar system, known as the Regenstrief Medical Record System (RMRS) [4], developed by Clement McDonald at the Regenstrief Institute, did improve patient care and needed to be disseminated.

Ove Wigertz subsequently came as a visiting professor to Salt Lake City in 1986. Among other things, Wigertz worked with the team at LDS Hospital to polish rules for predicting radiology findings [5,6]. Wigertz and the team saw first hand the challenges such as time, expertise, and data availability in the proper format for executing rules, and they became convinced that each hospital and clinical setting could not afford to format from scratch all of the medical knowledge that would be of assistance inpatient care [7–9].

Paul Clayton subsequently moved to what was then known as Columbia-Presbyterian Medical Center in New York in 1987, where he assembled a young team and emulated the HELP architecture with newer technology and database concepts. The issue of knowledge representation remained. One of the team members, George Hripcsak, was tasked with building the "rules engine" and used the experience of HELP and RMRS to guide his efforts. Clayton had received funding from IBM and had argued with them about the liability that might be incurred if the clinical information system gave

* Corresponding author at: Department of Biomedical Informatics, Columbia University, 622 W 168th Street, PH20, New York, NY 10032, USA. Tel.: +1 212 305 5712.
E-mail address: hripcsak@columbia.edu (G. Hripcsak).

https://doi.org/10.1016/j.artmed.2015.05.006
0933-3657/

wrong suggestions or reminders. He felt that the knowledge base should be public (as in any medical textbook) and then individuals or providers could choose to validate and use that knowledge rather than having it “baked into” a proprietary information system. The group identified the obstacles to sharing [10] and proposed that a common structure should be developed that would accommodate the logic of existing knowledge bases that are represented in a modular fashion [10].

Paul Clayton, Ove Wigertz, and Allan Pryor called for a retreat [11] to discuss the sharing of knowledge-based systems. Funding was obtained from IBM, the CAMDAT Foundation, and the law firm of De Forest and Duer, and the retreat was held at the Arden Homestead, a delightful rural setting in Harriman, NY that was donated to Columbia University by the Harriman family. In preparation for the retreat, George Hripcsak created a straw-man syntax [11] for the group to consider, derived primarily from HELP [12] with features from RMRS [4] such as its aggregation operators and the manner in which the result of a database query can be manipulated.

The retreat took place June 16–18, 1989 [11] with 18 informaticians from 10 institutions. The discussion was broad and covered all aspects of knowledge sharing [11], including the use of procedural versus declarative knowledge, vocabulary issues, validation, and liability. An extensive discussion ensued on where to focus initial sharing efforts: modular, independent knowledge bases like HELP and RMRS, in which each element of the knowledge base acts independently of the others, as happens with alerts and reminders; or modular, interdependent systems, such as those with disease profiles like QMR [13] and DXplain [14], in which sharing would be more difficult and require more research; or systems with tight interdependencies like CASNET [15].

With a majority vote of the participants, Clayton, Wigertz, Pryor, and Hripcsak pushed forward on the syntax for modular, independent knowledge bases. The individual elements of the knowledge base were named Medical Logic Modules (MLMs) [11,16], and the syntax was named the Arden Syntax [17] after the conference center. Its design goals were that the language should be easily readable, unambiguous, relatively easily written, and executable. It was noted that the ability to read the syntax—so that clinicians could verify the logic—was an even higher priority than the ability to write it, which might still require the help of knowledge engineers. The initial decision to base it on HELP and RMRS, which both had demonstrated success and were both well established in their sponsoring institutions, was maintained.

A number of design decisions were made before and after the retreat. The first decision was to organize an MLM as a set of slots where the knowledge-related slots were populated with a procedural programming language like those of HELP and RMRS, which were well-suited to the short alerts and reminders that were encoded in those languages. Sets of data would be manipulated within MLMs, and a concise, readable notation for that manipulation was desired. Loops and goto constructs from lower-level languages were rejected as uninterpretable by clinicians, and nested syntactic structures like those available in Lisp were also rejected. In the end, the data manipulation functions were borrowed from the functional programming paradigm, and more specifically from a precursor of that paradigm, the language APL [18]. While APL is infamous for its cryptic, unreadable code, APL’s functional constructs were well-suited to array processing. Constructs from HELP and RMRS were mapped to APL array constructs (e.g., Arden “any” is APL or-reduce) and given more readable names.

Arden, like APL, employed dynamic typing, avoiding type declarations and providing real-time type conversion, but expanded the types (Boolean, number, string, time, duration) and added a null. This approached was deemed most readable by clinicians. Lists were heterogeneous to support databases that could return mixed results, and two-dimensional arrays were supported implicitly through assignment into several lists. Primary time, defined generally as whatever time the institution deemed most important and set implicitly within the queries, was supported. Arden was purposely not given constructs that would make it a general-purpose programming language (e.g., goto), but through nesting of MLMs, general purpose algorithms could be implemented at the great expense of readability. The syntax of the language was verified using lex and yacc to find ambiguities, and then further evaluated by experts in computer languages, such as Ludemann [19]. The institution-specific component of data queries was separated from the sharable part of the MLM within curly brackets [20]. In a study of the sharability of MLMs, the institution-specific query was found to be the primary stumbling block [21]. Thus this issue came to be known as the curly braces problem, and it includes challenges at several levels: the terminologies used to encode clinical data elements, the data schema that contains the elements, the syntax used to query the schema, and the local characteristics of the data (availability, accuracy, etc.). A full treatment of these issues is beyond the scope of this paper, but progress has been made on several fronts (standards, web services, virtual health records). A number of other issues remained outside of the syntax but still important for implementation, such as synchronous versus asynchronous MLMs and how to evoke and inactivate MLMs.

Clement McDonald guided the developers to put Arden through the standards development process of the American Society for Testing and Materials (ASTM), a consensus standard organization. It was formally approved and published in 1992 [22].

During the period from 1989 to 1993, a large number of researchers and engineers worked on the syntax. Much of the documentation from that period is lost, but rather than name no one, we will name those for whom we have found records and we apologize to the many more whom we are missing. In addition to those named above, Johan van der Lei and Dick van Nes in particular worked with Hripcsak in the early days verifying the syntax in its formative stages. The syntax was first implemented at Columbia University Medical Center using a compiler and a pseudo-code interpreter [20]. Other early implementations—by 1990—of Arden included Erasmus’s by van der Lei and van Nes, Wigertz group’s including Magyar [23], LDS Hospital’s by Pryor and Peter Haug, and one by Doug McNair. A wide range of languages was used, including C++, PL/I, Prolog, Smalltalk, and APL.

The 1991 “Arden Syntax Task Group Membership” included Johan van der Lei, Peter Haug, Roger Strube, Clement McDonald, Doug McNair, Roger Brittain, Steward McMorran, and George Hripcsak. The 1991 “Arden Syntax Task Group on Validation/Verification” included Roger Brittain, James Cimino, John Edward Deaver, Charles Friedman, Peter Haug, George Hripcsak, Michael Kahn, Allan Pryor, Alan Rector, Ove Wigertz, Jeremy Wyatt, Shamsul Chowdhury, and Nunzia Giuse. Ludemann put in endless hours assessing and improving the syntax. We also note Hans Ahlfeldt, Ankica Babic, Carol Broverman, James Campbell, Bo Johansson, Stephen Johnson, Gil Kuperman, Ulli Prokosch, Terry Rankin, Harm Scherpbier, Soumitra Sengupta, Nosrat Shahsavar, Rick Spates, Walter Sujansky, Michael Swhe, and Gao Xiao-ming. It is something of a who’s who list of biomedical informatics. And finally, we note Nancy Roderer, who ran the MLM library at Columbia. The syntax and implementations were strong, but actual sharing remained light. Roderer famously quipped, “We got that MLM, and we are managing it extremely well”; the library served more as a proof of concept than an actual cache of knowledge, and other advances in addition to the syntax were necessary for sharing. Nevertheless, by 1994, eight institutions were using the syntax, and seven vendors had products or plans to support the syntax [24].

The early developers of Arden are indebted to those who followed and brought Arden to fulfill its promise. We look forward to many years of fruitful use and improvement.

References

[1] Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. J Med Syst 1983;7(2):87–102.
[2] Dudeck J, Clayton PD, Sebald P, Schmuecker P, Alexander R, Heeg M, et al. Erfahrungen bei der Anpassung des Krankenhaus-Informationssystems HELP an die Deutsche Umgebung. In: Hagenbuch RE, editor. Medizinishe Informatic in der Schweiz. Basel: Schwabe & Co; 1986. p. 127–46.
[3] Dudeck J, Sebald P, Prokosch HU, Michel A, Heeg M. The transfer of the hospital information system HELP into the environment of German hospitals. In: Proceedings of medical informatics Europeí87, vol. 1. 1987. p. 18–24.
[4] McDonald CJ. Action-oriented decisions in ambulatory medicine. Chicago: Year Book Medical Publishers; 1981.
[5] Clayton PD, Ostler DB, Gennaro JL, Beatty SS, Frederick PR. A radiology reporting system based on most likely diagnoses. Comp and Biomed Res 1980;13:258–70.
[6] Wigertz OB, Wigertz G, Clayton PD, Wong T, Tocino I, Morrison WJ. Predicting X-ray findings on the Utah-HELP system. In: Piene U, Kvernes S, editors. Proceedings of VII Nordic meeting on med and biolog engineering. Norway: Trondheim; 1987. p. 145–6.
[7] Wigertz OB, Clayton PD, Haug PJ, Pryor TA. Design of knowledge based systems for multiple use, truth maintenance and knowledge transfer. In: Proceedings of medical informatics Europe'87, vol. 2. 1987. p. 986–91.
[8] Wigertz OB, Pryor TA, Clayton PD, Haug PJ. Transportability of knowledge between expert systems. In: Hansen R, Solheim BG, O'Moore RR, Roger FH, editors. Proceedings of medical informatics Europe'88. 1988. p. 690–4.
[9] Wigertz OB, Clayton PD, Haug PJ, Pryor TA. Expert systems knowledge transfer. In: Reinhoff O, Piccolo V, Schneider B, editors. Expert systems and decision support in medicine. Hanover: Springer Verlag; 1988. p. 371–7.
[10] Clayton PD, Pryor TA, Wigertz OB, Johnson SB, Hripcsak GM. Sharing medical knowledge for automated decision-making. In: Greenes RA, editor. Proceedings of the 12th symposium on computer applications in medical care. Washington, DC: IEEE Computer Society Press; 1988. p. 591–4.
[11] Clayton PD, Pryor TA, Wigertz OB, Hripcsak G. Issues and structures for sharing knowledge among decision-making systems: the 1989 Arden homestead retreat. In: Kingsland LC, editor. Proceedings of the 13th symposium on computer applications in medical care. Washington, DC: IEEE Computer Society Press; 1989. p. 116–21.
[12] HELP Frame Manual, Version 1.6. Salt Lake City, UT: LDS Hospital; 1989.
[13] Giuse NB, Giuse DA, Miller RA. Computer assisted multi-center creation of medical knowledge bases. In: Greenes RA, editor. Proceedings of the 12th symposium on computer applications in medical care. Washington, DC: IEEE Computer Society Press; 1988. p. 509–83.
[14] Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain: an evolving diagnostic decision-support system. J Am Med Assoc 1987;258:67–74.
[15] Weiss S, Kulikowski C, Safir A. Glaucoma consultation by computer. Comp Biol Med 1978;8:25–40.
[16] Wigertz OB, Clayton PD, Hripcsak G, Linnarsson R. Knowledge representation and data model to support medical knowledge base transportability. In: Talmon J, Spiegelhalter J, editors. System engineering in medicine. Heidelberg: Springer Verlag; 1989. p. 80–90.
[17] Hripcsak G, Clayton PD, Pryor TA, Haug P, Wigertz OB, van der Lei J. The Arden syntax for medical logic modules. In: Miller Ra, editor. Proceedings of the 14th symposium on computer applications in medical care. Washington, DC: IEEE Computer Society Press; 1990. p. 200–4.
[18] Iverson KE. A programming language. New York: Wiley; 1962.
[19] Hripcsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD. Rationale for the Arden syntax. Comput Biomed Res 1994;27:291–324.
[20] Hripcsak G, Cimino JJ, Johnson SB, Clayton PD. The Columbia-Presbyterian Medical Center decision-support system as a model for implementing the Arden syntax. In: Proceedings of the 15th symposium on computer applications in medical care. New York: McGraw-Hill; 1991. p. 248–52.
[21] Pryor TA, Hripcsak G. Sharing MLM's: an experiment between Columbia-Presbyterian and LDS Hospital. In: Safran C, editor. Proceedings of the 17th annual symposium on computer applications in medical care. New York: McGraw-Hill; 1993. p. 399–403.
[22] American Society for Testing and Materials. E 1460 standard specification for defining and sharing modular health knowledge bases (Arden syntax for medical logic modules). 1992 Annual book of ASTM standards, vol. 14.01. Philadelphia: American Society for Testing and Materials; 1992. p. 539–87.
[23] Magyar G, Arkad K, Ericsson K-E, Gill H, Linnarsson R, Wigertz O. Realizing medical knowledge in MLM form as working modules in a patient information system. In: Timmers T, Blum BI, editors. Proceedings of the IMIA working conference on software engineering in medical informatics, Amsterdam, 8–10 October 1990. Amsterdam: North-Holland Press; 1990.
[24] Hripcsak G. The Arden syntax for medical logic modules: introduction. Comput Biol Med 1994;24:329–30.