

JSON-ASR: A lightweight data storage and exchange format for automatic systematic reviews of TCM

Ji Xu¹, Hongyong Deng ^{1*}

¹ Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China.

***Correspondence to:** Hongyong Deng, denghy@shutcm.edu.cn

Highlights

A JSON-based data format, JSON-ASR, was developed to store and exchange TCM clinical trial information for automatic systematic reviews. The JSON-ASR with features as follow:

- (1) Includes six sections of general, source, clinic, study, group, and outcome;
- (2) Lightweight on file size;
- (3) Plain-text format is easy to read and parse;
- (4) Nested tree structure ensures efficiency in storing and transmitting clinical data;
- (5) Excellent extensibility and compatibility.

Abstract

Objectives: The aim of this study was to investigate and develop a data storage and exchange format for the process of automatic systematic reviews (ASR) of traditional Chinese medicine (TCM).

Methods: A lightweight and commonly used data format, namely, JavaScript Object Notation (JSON), was introduced in this study. We designed a fully described data structure to collect TCM clinical trial information based on the JSON syntax.

Results: A smart and powerful data format, JSON-ASR, was developed. JSON-ASR uses a plain-text data format in the form of key/value pairs and consists of six sections and more than 80 preset pairs. JSON-ASR adopts extensible structured arrays to support the situations of multi-groups and multi-outcomes.

Conclusion: JSON-ASR has the characteristics of light weight, flexibility, and good scalability, which is suitable for the complex data of clinical evidence.

Keywords: Data storage and exchange, Automatic systematic reviews, Traditional Chinese medicine, JavaScript object notation

Abbreviations: ASR, Automatic Systematic Reviews; TCM, Traditional Chinese Medicine; JSON, JavaScript Object Notation; EBM, Evidence-Based Medicine; ICASR, The International Collaboration for Automation of Systematic Reviews; HL7, Health Level 7; CDISC, The Clinical Data Interchange Standards Consortium; XML, Extensible Markup Language.

Funding: This work was supported by the National Key R&D Program of China (Grant no. 2019YFC1709803) and National Natural Science Foundation of China (Grant no. 81873183).

Competing interests: The authors declare that there is no conflict of interests regarding the publication of this paper.

Citation: Xu J, Deng HY. SON-ASR: A lightweight data storage and exchange format for automatic systematic reviews of TCM. TMR Modern Herbal Medicine. 2021, 4(2):12.

Executive Editor: Chaoyong Wu

Submitted: 1 April 2021, **Accepted:** 21 April 2021, **Online:** 25 April 2021.

© 2021 By Authors. Published by TMR Publishing Group Limited. This is an open access article under the CC-BY license (<http://creativecommons.org/licenses/by/4.0/>).

Background

In the past 30 years, evidence-based medicine (EBM) has made significant developments from its establishment to maturity and has had a large impact on the model of health decision-making. However, EBM still has many problems, in which the most criticized are the long procedure cycle and low timeliness of classic systematic reviews [1]. Researchers have attempted to find solutions, such as automation of systematic reviews (ASR), to address this issue. The International Collaboration for Automation of Systematic Reviews (ICASR) pointed out in 2018 [2] that without an automated method for reviewing thousands of research articles, including the many published every year, findings might be overlooked when developing new policies; automated tools for systematic review would enable more transparent and timely reviews, thus maximizing the potential for identifying and translating research findings into practical applications. The research of ASR has been continuously enriched, and now, there are many informatic technologies, data storage and exchange mechanisms, algorithms, and visualization techniques involved in this field. Among them, an efficient data storage and exchange format is essential. This is beneficial for smooth collaboration among different working groups, different system platforms, different systematic reviews software, and different functional modules of evidence production tools.

In the traditional process of system review, printed paper forms and/or Microsoft Excel spreadsheets are usually used for data collection and storage. Cochrane had recommended its data sheet for making Cochrane SR [3], but this kind of table is not suitable for automatic processing by computers. There are currently no public reports on ASR data storage and exchange standards. For this study, we referred some resources of clinical trials, such as Health Level 7 (HL7) [4] and the Clinical Data Interchange Standards Consortium (CDISC) [5]. We also considered the possibility of using different file formats in ASR data. In terms of maturity and universality, Extensible Markup Language (XML) [6,7] and JavaScript Object Notation (JSON) have obvious advantages, and JSON is better at lightweight on its size [8].

JSON is a flexible, efficient, and open standard file format for data interchange. It is based on a subset of the JavaScript Programming Language Standard ECMA-262 3rd Edition - December 1999. JSON is a human-readable text format that is completely language-independent but uses conventions that are familiar to programmers of the C-family of languages, including C, Java, JavaScript, Python, and many others

Submit a manuscript: <https://www.tmrjournals.com/mhm>

[9,10]. These properties make JSON an ideal data-interchange format for ASR.

Methods

JSON Syntax

JSON is a language-independent data format. It was derived from JavaScript, but many modern programming languages include code to generate and parse JSON-format data. The official filename extension for JSON is “.json”. JSON is built on two structures: a collection of key (name)/value pairs, and an ordered list of values. The former is also known as an object. An object is an unordered set of key/value pairs. An object begins with “{” (left brace) and ends with “}” (right brace). Each key is followed by “:” (colon) and the key/value pairs are separated by “,” (comma). In this study, we designed the framework and elements of JSON-ASR according to the characteristics of ASR data while following the JSON syntax.

JSON-ASR Sections

A JSON-ASR file is a predefined dataset to describe a clinical evidence object. Different parts of clinical evidence can be mapped to different sections of JSON files. Clinical evidence usually consists of six parts: general information, source of evidence, clinical data, trial design, grouping, and outcomes. Each part corresponds to a sub-object (section) under the main JSON object. For those multi-member sections, such as sources, groups, and outcomes, arrays of objects were employed rather than a single object.

Key/Value Pairs

JSON uses key/value pairs to describe entity information. The key is a string in double quotes to name the value. The values vary in types such as string, number, array, Boolean, and object. A string-type value should be double quoted, while a number, or Boolean (i.e., true or false) value without quotes. An array begins with “[” (left bracket) and ends with “]” (right bracket); elements in an array are separated by “,” (comma). JSON object structures can be nested, which means the value of an object can be another object or even an array of objects.

Results

Framework of JSON-ASR

JSON-ASR is a plain-text file format that is easy to read and parse. Each JSON-ASR file (*.json) stores the entire clinical evidence data in a nested object structure. There are six sections, namely, general, source, clinic, study, group, and outcome, in the main object of JSON_AS (see [Figure 1](#)). Each section has a

key/value pair structure to describe the information within its field. All keys are named and followed by an object (general, clinic, study sections) or an array of objects (source, group, and outcome sections). For example, the general information of a clinical evidence, such as unique id, name, create date, and editor, can be fully presented in just one object, but the group section requires a list of objects to record its grouping data. For a clinical trial, there are at least two groups, namely, intervention and control.

JSON-ASR allows up to three levels of nesting, for example, “source-authors-first author.” The evidence source section can clearly inform the first author who reported the trial, and the path of “outcome-treatments-data” can be followed to retrieve the exact group’s outcome value. This nested tree structure ensures efficiency of JSON-ASR in storing and transmitting clinical evidence data.

Key/value pairs

JSON-ASR describes data details in the form of key-value pairs; some of them are required and others are optional. In this paper, the value types are limited to string, number, array, and object (see Table 1). There are about 80 to 140 pairs in a typical JSON-ASR file; the actual number is based on the number of sources, groups, and outcomes within it. Accordingly, the size of a JSON-ASR file is approximately 9 kilobytes, with a range from 4 kilobytes to 15 kilobytes.

In the general section, there are six pairs for recording evidence and normal information, and the source section has eight pairs. Two source-section elements, namely, journal and authors, have nested objects as value. The former has an object of journal, such as

```
"journal": {
  "name": "山西中医",
  "language": "中文",
  "year": "2020",
  "volume": "36",
  "issue": "1",
  "pages": "12-14"
},
```

and the authors’ object is

```
"authors": {
  "first author": "李坤",
  "respondence": " ",
  "all authors": "李坤",
  "organization": "江苏省靖江市中医院",
  "contact": " "
}.
```

Similarly, the study object has a sub-object of risk of bias, which provides full details of the Cochrane seven risk factors:

```
{
  "general": {
    ... (key/value pairs of general information)
  },
  "source": [
    {
      ... (key/value pairs of source data)
    },
    {
      ... (key/value pairs of source data)
    },
    ... source object n
  ],
  "clinic": {
    ... (key/value pairs of clinical data)
  },
  "study": {
    ... (key/value pairs of study data)
  },
  "group": [
    {
      ... (key/value pairs of group data)
    },
    {
      ... (key/value pairs of group data)
    },
    ... group object n
  ],
  "outcome": [
    {
      ... (key/value pairs of outcome data)
    },
    {
      ... (key/value pairs of outcome data)
    },
    ... outcome object n
  ]
}
```

Figure 1. Framework of JSON-ASR

```
"rob": {
  "randomize": "低风险",
  "randomize desc": "按照随机数字表法分为两组各42例。(p13)",
  "allocation": "不清楚",
  "allocation desc": "本实验未说明分配隐藏方法.",
  "act blind": "高风险",
  "act_blind desc": "文章未说明受试者和操作者的盲法.",
  "test blind": "高风险",
  "test_blind desc": "本实验未说明检测者盲法.",
  "incomplete": "低风险",
  "incomplete desc": "本实验没有数据脱落.",
  "selective": "低风险",
  "selective desc": "方法学涉及的结局指标均被报道.",
  "other": "低风险",
  "other desc": "无其他明显的偏倚."
}.
```

Table 1. Main key/value pairs of JSON-ASR (Level 1,2)

Code	Key	Value Type	Required	Sample Value
1	general	object	yes	
1.1	id	string	yes	"2020103001003"
1.2	status	string	yes	"已审核"
1.3	create date	string	no	"2020-10-30"
1.4	name	string	yes	"李坤 2020"
1.5	editor	string	yes	
1.6	checker	string	yes	
2	source	array of object	yes	
2.1	type	string	yes	"期刊文献"
2.2	title	string	yes	
2.3	journal	object	yes	
2.4	authors	object	yes	
2.5	funds	string	no	
2.6	doi	string	no	"1000-7156(2020)01-0012-03"
2.7	abstract	string	no	
2.8	full text	string	no	"pdf/2020103001003.pdf"
3	clinic	object	yes	
3.1	TCM disease	string	yes	"胃痞病"
3.2	TCM pattern	string	no	"脾胃虚弱证"
3.3	WM disease	string	yes	"化疗性胃肠道反应"
3.4	sample size	number	yes	80
3.5	age	string	no	
3.6	history	string	no	
3.7	include criteria	string	no	
3.8	exclude criteria	string	no	
4	study	object	yes	
4.1	type	string	yes	"随机对照试验"
4.2	blind	string	yes	"双盲"
4.3	duration	string	yes	"4周"
4.4	setting	string	no	"住院"
4.5	risk of bias	object	yes	
5	group	array of object	yes	
5.1	type	string	yes	"观察组"
5.2	name	string	yes	"中药组"
5.3	sample size	number	yes	40
5.4	intervention	object	yes	
6	outcome	array of object	yes	
6.1	name	string	yes	"中医证候积分"
6.2	follow up	string	yes	"4周"
6.3	direction	string	yes	"bigger is better"
6.4	endpoint	string	yes	"终点值"
6.5	type	string	yes	"连续型数据"
6.6	data	array of object	yes	

An intervention object is implemented in the group section, which includes drug, administration, and dosage information:

```
"intervention": {
  "name": "和胃消痞汤",
  "form": "汤剂",
  "admin": "口服",
```

```
"dosage": "75ml",
"frequency": "2天",
"duration": "4周",
```

```
}.
```

For each outcome, there are always two or more groups' test values to be reported. Hence, an array is

needed in the outcome section to present all groups' data:

```
"outcome data": [
  {
    "group name": "中药组",
    "data": "7.84,5.84,42"
  },
  {
    "group name": "西药组",
    "data": "11.54,6.12,42"
  }
]
```

The value syntax of “data” is different based on its own type. For continuous data, there are three numbers separated by a comma, namely, *mean*, *standard deviation*, and *sample size*; for dichotomous data, only *events number* and *sample size* are listed.

Discussion

The JSON data format is widely used in data storage and exchange [11,12], which makes JSON-ASR have the most compatibility. A JSON-ASR file organizes clinical evidence information in the form of objects and presents it in plain-text format, which is not only suitable for computer processing but also convenient for human reading. The symbol-based key/value pair system can clearly describe the data and minimize redundant information, thus ensuring a lightweight JSON-ASR file. We compared 243 trials' data sizes of different file types. The size of the original PDF files ranged from 26 to 3066 kilobytes and averaged 844.26 kilobytes; XML files were smaller than PDFs, ranging from 9 to 57 and averaging 14.84 kilobytes; the JSON-ASR files were the smallest, ranging from 4 to 39 and averaging 9.3 kilobytes. Considering that the amount of information carried by the three types of files is almost the same for systematic reviewing, JSON-ASR has higher storage and transmission efficiency compared with the other file types.

JSON-ASR predefines about 80 elements (objects and key/value pairs) of clinical evidence, which is enough for normal systematic reviews. We also verified this on a TCM-ASR prototype system [13]. However, we still believe that TCM-ASR is a file format with excellent extensibility; hence, it is recommended to make appropriate extensions in the case of ensuring compatibility in actual use. As readers may have noticed, TCM-ASR is designed and best suited for randomized controlled trial (RCT) data, and it is not inherently suitable for all types of clinical studies, such as one-arm trials and diagnostic accuracy tests. Nevertheless, JSON-ASR is still a good and basic reference format for the other types of clinical evidence data.

Conclusion

This paper reports an attempt to establish a storage and exchange format that supports clinical evidence for automatic systematic reviews. JSON-ASR, a JSON-based lightweight and smart data format was developed. The JSON-ASR file includes six sections of general, source, clinic, study, group, and outcome; each has some key/value pairs to describe the details of evidence. The JSON-ASR file has nested objects and a tree structure to fit the complex data of clinical evidence. The size of a JSON-ASR file is smaller; hence, it has more advantages in storage and transmission. In conclusion, JSON-ASR is a suitable data format standard of clinical evidence that can effectively promote the development and application of automatic systematic reviews.

Data Availability

The data used to support the findings of this study are available from the corresponding author on request.

References

1. Beller E, Clark J, Tsafnat G, et al., On behalf of the founding members of the ICASR group. Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). *Systematic Reviews*, 2018, 7:77.
2. O'Connor AM, Tsafnat G, Gilbert SB, et al. Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR). *Systematic Reviews*, 2018,7:1-5.
3. Higgins JPT, Thomas J, Chandler J, et al [Internet]. *Cochrane handbook for systematic reviews of interventions version 6.2 (updated February 2021)* [cited 2021 March 25]. Available from www.training.cochrane.org/handbook.
4. HL7 International [Internet]. About HL7 [cited 2021 March 25]. Available from www.hl7.org.
5. The Clinical Data Interchange Standards Consortium [Internet]. CDISC [cited 2021 March 25]. Available from www.cdisc.org.
6. W3C [Internet]. Extensible Markup Language (XML) [cited 2021 March 25]. Available from www.w3.org/XML.
7. Bond RR, Finlay DD, Nugent CD, et al. Research article XML-BSPM: an XML format for storing Body Surface Potential Map recordings. *BMC Medical Informatics and Decision Making*, 2010, 10:28

8. Json.org [Internet]. Introducing JSON [cited 2021 March 25]. Available from www.json.org/json-en.html.
9. ECMA [Internet]. Standard ECMA-404: The JSON Data Interchange Syntax 2nd edition [cited 2021 March 25]. Available from www.ecma-international.org/publications/standards/Ecma-404.htm.
10. Wikipedia [Internet]. JSON [cited 2021 March 25]. Available from en.wikipedia.org/wiki/JSON.
11. Srivastava DJ, Vosegaard T, Massiot D, et al. Core Scientific Dataset Model: A lightweight and portable model and file format for multi-dimensional scientific data. PLoS ONE. 2020, 15(1): e0225953.
12. Kobayashi N, Ishii M, Takahashi S, et al. Semantic-JSON: a lightweight web service interface for Semantic Web contents integrating multiple life science databases. Nucleic Acids Research, 2011, 39:W533–W540.
13. Hongyong Deng [Internet]. Automatic Systematic Review Platform for TCM Clinical Evidences Ver 1.01 [cited 2021 March 25]. Available from 124.70.215.69.