

## Datos estructurados

Son un formato estandarizado usado por google para poder proporcionar información de una página y poder clasificarla. El formato recomendado de este es JSON y se guarda en la etiqueta `script[type="application/ld+json"]`

```
<script type="application/ld+json">
{
  "@context": "https://schema.org/",
  "@type": "Recipe",
  "name": "Party Coffee Cake",
  "author": {
    "@type": "Person",
    "name": "Mary Stone"
  },
  "datePublished": "2018-03-10",
  "description": "This coffee cake is awesome and perfect for parties.",
  "prepTime": "PT20M"
}
</script>
```

Fuente: [Google](#)

Este JSON lo encontramos en las descripciones de los jobs haciendo la búsqueda de la etiqueta `script[type="application/ld+json"]` dentro del HTML.

[Welcome page](#)[Log back in!](#)

### Content Editor, Evidence-Based Nursing

UK-London

Job ID	2021-1281	UK-London	Salary	£36,216 pro-rata	Post End Date	01/04/2021
--------	-----------	-----------	--------	------------------	---------------	------------

#### Overview

**Position:** Content Editor, Evidence-Based Nursing  
**Contract:** Part-time - 1 year Fixed term contract  
**Salary:** from £36,216 per annum (pro rata)  
**Start date:** ASAP

BMJ is the innovative publishing division of the British Medical Association and one of the world leaders in medical publishing with a strong brand recognised internationally. One of the key product areas, BMJ Journals, produces over 70 specialist clinical medical titles, many of which are leaders in their fields.

#### The role

BMJ currently has a vacancy for a part-time Content Editor to work on *Evidence-Based Nursing*.

```
<!-- end google tag manager -->
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "JobPosting",
  "name": "Content Editor, Evidence-Based Nursing",
  "description": "BMJ is the innovative publishing division of the British Medical Association and one of the world leaders in medical publishing with a strong brand recognised internationally. One of the key product areas, BMJ Journals, produces over 70 specialist clinical medical titles, many of which are leaders in their fields.",
  "datePublished": "2021-03-10",
  "author": {
    "@type": "Organization",
    "name": "British Medical Association",
    "url": "https://www.bma.org.uk",
    "logo": "https://www.bma.org.uk/logo.png"
  },
  "jobLocation": {
    "@type": "Location",
    "address": {
      "@type": "PostalAddress",
      "streetAddress": "11 St Andrews Place, Regents Park",
      "addressLocality": "London",
      "postalCode": "NW1 4NU",
      "addressCountry": "UK"
    }
  },
  "contractType": "Fixed term",
  "employmentType": "Part-time",
  "salary": {
    "@type": "Salary",
    "value": "£36,216 per annum (pro rata)"
  },
  "startDate": "2021-04-01",
  "endDate": "2022-03-31"
}
```

Estos datos estructurados los podemos encontrar en gran variedad de jobsites (Uno de los más comunes que nos topamos es icims) y aquí por lo general podremos encontrar el “dateposted” y “dateclosed” (para los date closed por lo general aparecen en el JSON como validThrough”) pero también en otros casos podemos encontrar el location, Salary y Jobtype.

```

<script type="application/ld+json">
{
  "@context" : "http://schema.org/",
  "@type" : "JobPosting",
  "title" : "Production Operative - 4 on 4 off",
  "description" : "<p> Ambitek are currently looking to recruit a Engineering
Operative for our client based in Haydock.</p><p>The company specialise in the manufacture
of plastic products, from medical supplies, catering supplies etc.</p><p>Originally,
starting out as a end of line operator, you will be expected to pack products for despatch,
however the company are looking for more than packers, they are looking for candidates
eager to learn and progress, eventually moving onto fault finding machinery.</p><p>The role
is on rotating days and nights, 4 on 4 off, 12 hour shifts.<br />You will be give 2 x 15
minute paid breaks / 1 x 30 minute unpaid break<br />You will also be given 20 days
holiday</p><p>Rate of pay is £8.85 - This is flat, no enhancement for OT, night shift
etc.</p><p>If you are interested, please send your CV in for consideration,</p>",
  "identifier": {
    "@type": "PropertyValue",
    "name": "Ambitek Limited",
    "value": "SJB212855686"
  },
  "baseSalary": {
    "@type": "MonetaryAmount",
    "currency": "GBP",
    "value": {
      "@type": "QuantitativeValue",
      "minValue": 8.84,
      "maxValue": 8.85,
      "unitText": "YEAR"
    }
  },
  "jobLocation": {
    "@type": "Place",
    "address": {
      "@type": "PostalAddress",
      "addressCountry": "UK"
    }
  },
  "employmentType": "",
  "datePosted" : "20210315 11:47:06",
  "validThrough" : "20210412 11:47:06",
  "hiringOrganization" : {
    "@type" : "Organization",
    "name" : "Ambitek Limited"
  }
}
</script>

```

La forma de poder extraer dicha información es usando este fragmento de código en el Step jobdescription.

```

// Step: Jobdata
if(document.querySelector('script[type="application/ld+json"]')){
  // Extract text on the script
  var html = document.querySelector('script[type="application/ld+json"]').textContent.trim().replace(/\s+/g, ' ').replace(/\\@/gi, "");
  // convert text to json
  var json = JSON.parse(html);
  /*var html = document.querySelector('script[type="application/ld+json"]').textContent.trim().replace(/\s+/g, ' ').replace(/\\@/gi, "");
  var json = JSON.parse(html);*/
  var date = json.graph[1].datePublished.split("T").shift().split("-");
  job.dateposted_raw = date[1]+"/"+date[2]+"/"+date[0];
}

```

Este lo que hace es extraer el contenido que hay dentro de la etiqueta `script[type="application/ld+json"]` y limpiar todo lo que vuelve invalido el JSON

(Enter “\n”, @, etc), despues de extraerlo convierte la información extraída en un objeto JSON, haciendo uso de la función “JSON.parse()”.

## Example - Parsing JSON

Imagine we received this text from a web server:

```
{ "name": "John", "age": 30, "city": "New York" }
```

Use the JavaScript function `JSON.parse()` to convert text into a JavaScript object:

```
var obj = JSON.parse('{ "name": "John", "age": 30, "city": "New York" }');
```

Fuente: [W3schools](https://www.w3schools.com/js/json_parse.asp)

Ejemplo práctico:

Caso ICIMS

1) Hacemos la extracción del JSON.

```
> var iframe = 'iframe[id="icims_content_iframe"]';
var iframedocument = document.querySelector(iframe).contentWindow.document;
var html =
iframedocument.querySelector('script[type="application/ld+json"]').textContent.trim().replace(/\s+/g, '
').replace(/\\@/gi, "");
// convert text to json
var json = JSON.parse(html);
< undefined
> json
< {hiringOrganization: {...}, validThrough: "2021-03-31T23:00:00.000Z", jobLocation: Array(1), employmentType: "OT
HER", type: "JobPosting", ...}
  context: "http://schema.org"
  datePosted: "2019-03-19T15:02:01.925Z"
  description: "<h2>Overview</h2><p><strong>Position:</strong>&nbsp;<strong>Content Editor, Evidence-Based Nursing</p>
employmentType: "OTHER"
  hiringOrganization: {type: "Organization", name: "UNAVAILABLE", sameAs: "UNAVAILABLE"}
  jobLocation: Array(1)
    0:
      address:
        addressCountry: "UK"
        addressLocality: "London"
        addressRegion: "UNAVAILABLE"
        postOfficeBoxNumber: "UNAVAILABLE"
        postalCode: "WC1H 9JR"
        streetAddress: "UNAVAILABLE"
        type: "PostalAddress"
      __proto__: Object
      type: "Place"
      __proto__: Object
      length: 1
      __proto__: Array(0)
      title: "Content Editor, Evidence-Based Nursing"
      type: "JobPosting"
      url: "https://careers-bmj.icims.com/jobs/1281/content-editor%2c-evidence-based-nursing/job"
      validThrough: "2021-03-31T23:00:00.000Z"
      __proto__: Object
```



2) Extraemos la información que necesitamos.

```
> var date = json.datePosted.split("T").shift().split("-");
  console.log("Date posted: "); dateposted_raw = date[1]+"/"+date[2]+"/"+date[0];

Date posted:
< "03/19/2019"

> var datec = json.validThrough.split("T").shift().split("-");
  console.log("Date closed: "); dateclosed_raw = datec[1]+"/"+datec[2]+"/"+datec[0];

Date closed:
< "03/31/2021"

location = json.jobLocation[0].address.addressLocality + ", " + json.jobLocation[0].address.addressCountry;
> POST https://m.addthis.com/live/red_lojson/100eng.json?sh=0&ph=2990&iyh=2899&dt=_Dfalse%26j_addthis_widget.js:70
an1offset%3D-300%26jun1offset%3D-300%26in_iframe%3D1&pfm=0&icns= net::ERR_BLOCKED_BY_ADBLOCKER
"London, UK"
```

Y como vemos en el job como tal, no tenemos la fecha de publicación, pero la pudimos conseguir en el json, la fecha de cierre está un día atrasada pero es porque en el json aparece que cierra a las 23:00 del día anterior.

UK-London

**Job ID** 2021-1281    **UK-London**    **Salary** £36,216 pro-rata

**Post End Date** 01/04/2021

## Overview

**Position:** Content Editor, Evidence-Based Nursing

**Contract:** Part-time - 1 year Fixed term contract

**Salary:** from £36,216 per annum (pro rata)

**Start date:** ASAP

BMJ is the innovative publishing division of the British Medical Association and one of the world leaders in medical publishing with a strong brand recognised internationally. One of the key product areas, BMJ Journals, produces over 70 specialist clinical medical titles, many of which are leaders in their fields.

```
> var iframe = 'iframe[id="icims_content_iframe"]';
var iframedocument = document.querySelector(iframe).co
var html =
iframedocument.querySelector('script[type="application/
m().replace(/\s+/g, ' ').replace(/\@/gi, "");
// convert text to json
var json = JSON.parse(html);
< undefined

> json
< {hiringOrganization: {...}, validThrough: "2021-03-31T2
  Array(1), employmentType: "OTHER", type: "JobPosting"
    context: "http://schema.org"
    datePosted: "2019-03-19T15:08:12.412Z"
    description: "<h2>Overview</h2><p><strong>Position
    employmentType: "OTHER"
    ▶ hiringOrganization: {type: "Organization", name: "U
    ▶ jobLocation: [{...}]
    title: "Content Editor, Evidence-Based Nursing"
    type: "JobPosting"
    url: "https://careers-bmj.icims.com/jobs/1281/conte
    validThrough: "2021-03-31T23:00:00.000Z"
    ▶ __proto__: Object
      validThrough:
```