

Accessing time series

through a web service

according to the SDMX standard

v 2.2

Contents

Release history

N°	Date	Modifications
2.0	3/7/2017	
2.1	11/10/2017	A new fonctionnality : access to the previous versions of an amount of series
2.2	03/06/2020	Slight changes for release on the web portal api.insee.fr and minor corrections

Presentation of the service.....	3
The SDMX web service.....	3
About SDMX.....	4
I. Technical aspects.....	5
II. Simplified access to the series with their identifiers (idbank).....	5
Response message content.....	5
URL syntax.....	7
Response structure.....	8
Example.....	8
III. Access by dataflow.....	9
Response message content.....	9
URL syntax.....	11
Response structure.....	13
IV. Access to structural metadata.....	15
URL syntax.....	15
Response structure.....	15
Examples.....	16
V. Access to previous releases.....	17
Release calendar.....	18
Error handling.....	19

Presentation of the service

This REST web service allows to access in a structured way to the time series provided by the Insee web site. It meets the SDMX standards. This web service is available through two ways:

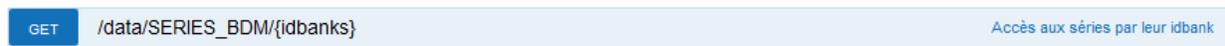
- directly from <https://bdm.insee.fr>. The service is only reachable through an HTTPS request ;
- by the Insee's API portal, available at <https://api.insee.fr>, with required sign in.

In the rest of this document, requests will start with <https://bdm.insee.fr/series/sdmx/> to enable a direct illustration with concrete examples. When using through the API portal, the instructions to be used are after those after this URL. For instance, when looking for a series through its identifier:

Direct HTTPS request

https://bdm.insee.fr/series/sdmx/data/SERIES_BDM/idbanks?optionalparameters

From the "Console de l'API" webpage on api.insee.fr



GET /data/SERIES_BDM/{idbanks} [Accès aux séries par leur idbank](#)

The series set includes various statistical indicators (short-term, structural, specific, ...) related to France or its regions and particularly time series. An indicator can be split into many basic series, corresponding to all possible crossings of variables. For example, 50 economic activities x 2 frequencies [monthly and quarterly] x 2 adjustments [raw and seasonally adjusted] = 200 different series.

Each time series contains the latest observation but also historical data, with a different anteriority depending on the nature of the indicators.

The SDMX web service

- allows to retrieve:
 - time series data
 - structural metadata describing the characteristics of the series
 - the release calendar of indicators
- is free of charge
- limits the amount of data that can be retrieved in a single request, so as not to overload the server and penalise other users

Data can be quickly requested using the [simplified access to the series with their identifiers](#). This feature requires to get the series identifiers (idbank) from INSEE website. It is recommended for retrieving a limited number of series and/or series from different thematic dataflows.

When the user wishes to extract detailed and well-structured data for an indicator, predefined thematic groups are available.

For some specific series (selected for their popularity, or concerning contract escalations), an access to the previous releases is provided.

About SDMX

The Statistical Data and Metadata eXchange (SDMX) initiative is sponsored by seven international organisations in order to provide standards for the exchange of statistical information.

The [SDMX 2.1](#) specification supplies a complete and accurate norm to structure statistical data and metadata, and a description of REST web services for accessing them: It will be respected as far as possible.

Eurostat provides a simple description of this standard <http://ec.europa.eu/eurostat/web/sdmx-web-services/sdmx> and comprehensive guides to learn how to use it:

- https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Main_Page

The implemented SDMX objects that can be collected through this web service are:

- **Dataset:** set of data, organised according to a predefined structure
- **Data Structure Definition (DSD):** metadata describing the structure and organisation of a dataset, the statistical concepts and the codelists used within it
- **Concept Scheme:** list of the concepts used by the DSDs
- **Codelist:** predefined list of codes (id + name) from which some statistical concepts take their values
- **Dimension:** identifies the time series in a dataset by combining concepts and codelists

- **Attribute:** gives additional information about the series or the observations, but do not affect the dataset structure itself
- **Dataflow:** structure for classifying the datasets
- **Category Scheme:** dataflows sorting tree

I. Technical aspects

The needed information can be retrieved from this web service with HTTPS GET requests only. The response message format will be SDMX-ML version 2.1, which depends on the requested resource:

- data: *StructureSpecificData* or *GenericData*
- structural metadata: *Structure*
- release calendar: *RSS 2.0*

These formats are briefly described in the next sections, detailed information is available in the standard technical specification ([section 3](#)).

When several formats are possible for a single resource, a choice can be made with content negotiation (using HTTP request header *Accept*).

HTTP response compression will be achieved if the client requests for it (header *Accept-Encoding: gzip*).

In case of client error (bad syntax for example) or server error (database unavailable for example), a SDMX response error message will be sent with a specific HTTP status code.

II. Simplified access to the series with their identifiers (idbank)

Response message content

Ten attributes describe the time series:

attribute id	description	format
IDBANK	numerical identifier	9 digits: 000639196

TITLE_FR	title in French	240 characters maximum
TITLE_EN	title in English	240 characters maximum
LAST_UPDATE	last update date	yyyy-mm-dd
UNIT_MEASURE	unit	12 characters maximum
UNIT_MULT	unit multiplier	integer between 0 and 9
REF_AREA	reference area	12 characters maximum
DECIMALS	decimals	integer between 0 and 9
BASE_PER	base period	2010 or 2008-T4 for example

The BASE_PER attribute is optional and is only displayed when not null. The nine other ones are always present in the response.

Seven characteristics describe the observations:

* Time period (marked as TIME_PERIOD) has a different format depending on the series frequency (YYYY stands for the year and n the period number) :

frequency	format	example	meaning
Annual	YYYY	2013	year 2013
Quarterly	YYYY-Qn	2014-Q3	third quarter of 2014
Monthly	YYYY-nn	1990-09	September 1990
Bi-monthly	YYYY-Bn	2010-B4	from July 1, 2010 to August 31, 2010
Semi-annual	YYYY-Sn	2012-S2	from July 1, 2012 to December 31, 2012

* The OBS_STATUS attribute is always present and explains the status of the observation, with different values:

- A: final data
- P: provisional
- SD: semi-final
- R: revised
- E: estimated
- O: missing

* The OBS_QUAL attribute is always present and explains the quality of the observation, with the following:

- E: estimated
- P: provisional
- SD: semi-final

- DEF: final data
- F : forecast

* The OBS_CONF attribute is optional and explains the confidentiality status of the data with the following values :

- C : confidential
- Q : covered by statistical confidentiality

* The OBS_REV attribute is optional and explains the status of the revisions with the following value :

- 1 : revision

* The DATE_JO attribute is optional, and added when an observation is released in the Official Journal of French Republic (journal-officiel.gouv.fr). It follows the format "yyyy-mm-dd", such as "2016-04-14".

Observation value

The observations are shown in anti-chronological order, the most recent ones listed first.

A dot is used as decimal mark, and there is no thousands separator: OBS_VALUE="1346.4" for example.

When the value is missing, OBS_VALUE="NaN" (« Not a Number ») is displayed.

URL syntax

The HTTP request for accessing the data through the series identifiers follows the pattern:

https://bdm.insee.fr/series/sdmx/data/SERIES_BDM/idbanks?optionalparameters

The *idbanks* part consists of a list of 9-digit identifiers separated by the « + » character, such as « 001565183+001690224+000067677 ».

It is not possible to retrieve more than 400 series in a single request.

The optional parameters are used to limit the amount of data sent back for each series. If they are not specified, all available values are displayed. The usable parameters are the following:

- **startPeriod**: beginning of the reporting period: 2001, 2010-Q3, 1990-09, ...
- **endPeriod**: end of the reporting period: 2014, 2013-S2, ...
- **firstNObservations**: integer N allowing to get the N oldest values of the series
- **lastNObservations**: integer N allowing to get the N most recent values of the series

The client can select the output format by content negotiation, using the *Accept* header value in his https request. The two possible values are the following:

- application/vnd.sdmx.structurespecificdata+xml;version=2.1: StructureSpecificData format
- application/vnd.sdmx.genericdata+xml;version=2.1: GenericData format

The StructureSpecificData format, more lightweight, is advised to access the series with their idbank. The GenericData format can also be requested, it is described in the next section [access to the series by group](#).

By default, the response will use the StructureSpecificData format.

Response structure

The SDMX message includes a header (XML element <Header>) before the data providing general information about the service.

In the StructureSpecificData format, each time series is presented in a <Series> element with multiple XML attributes corresponding to the idbank and the characteristics of the series. It has n children <Obs> with XML attributes TIME_PERIOD, OBS_VALUE, OBS_STATUS OBS_TYPE, OBS_QUAL and potentially OBS_REV, OBS_CONF and DATE_JO.

Example

The following URL provides data since 2010 for the 3 series 001565183, 001690224 and 000067677:

https://bdm.insee.fr/series/sdmx/data/SERIES_BDM/001565183+001690224+000067677?startPeriod=2010

III. Access by dataflow

In the SDMX information model, data are presented in a dataflow. Each time series is identified by crossing several dimensions, it can have many attributes and it contains a list of observations. These observations are composed of a value, a date and attributes.

Therefore, a dataflow is closely linked to a Data Structure Definition (DSD): this structure defines the dimensions and the mandatory or optional attributes used in the dataflow. The access to the different available structures (including the DSDs) is described in the next section.

It is not possible to retrieve more than 2000 series in a single request. Most of the dataflows have fewer than 2000 series. For large dataflows with more than 2000 series, the client will choose a criterion ([key part of the request](#)) to cut the dataflow.

For requests leading to more than 2000 series, the response is an error message: "this request is to big".

Response message content

For time series there is one DSD for each dataflow, it is formed with the following components:

- the dimensions are equivalent to the differentiation criteria
- the attributes of the series correspond to some of their characteristics
- the attributes of the observations are the status of the data (definitive, provisional, ...), its quality status, its type and potentially its revised or confidential status the embargo date and the released date in Journal Officiel

Dimensions

There are at least 6 dimensions for each dataflow. They match the lists of items used to select the desired series on the BDM website.

List of the six mandatory dimensions

FREQ	Frequency	1 letter: A, T, M, B or S
REF_AREA	Reference area	12 characters maximum
UNIT_MEASURE	unit	12 characters maximum
NATURE	Type of the indicator : index, variation, year-	12 characters maximum

	on-year	
INDICATEUR	measurement	12 characters maximum
CORRECTION	raw, seasonal adjustment, seasonal adjustment-working days,	12 characters maximum

And some optional dimensions depending on each dataflow.

Exemple : for the IPCH-2015 dataflow (harmonized consumer price indices) there are two complementary dimensions :

BASIND : basing year for the index

COICOP2016 : nomenclature of product

The identifiers and the possible values are described in the associated DSD.

Attributes of the time series

There is a maximum of 7 attributes at the series level. These attributes provide additional information to describe the time series. Among these 7 attributes, Base period is only displayed when appropriate and the 6 other ones are always present:

attribute id	description	format
IDBANK	numerical identifier previously used in the bdm.insee.fr web-site	9 digits: 000639196
FREQ	frequency	1 letter: A, T, M, B or S
TITLE_FR	title in French	240 characters maximum
TITLE_EN	title in English	240 characters maximum
LAST_UPDATE	last update date	yyyy-mm-dd
UNIT_MEASURE	unit	12 characters maximum
UNIT_MULT	unit multiplier	integer between 0 and 9
REF_AREA	reference area	12 characters maximum
DECIMALS	decimals	integer between 0 and 9
BASE_PER	base period	2010 or 2008-T4 for example

The BASE_PER attribute is optional and only displayed when not null. The nine other ones are always present in the response.

Attributes of the observations

identifiant	statut	format
TIME_PERIOD	mandatory	YYYY, YYYY-Qn,... see below
OBS_STATUS	mandatory	One or two letters from CL OBS STATUS
OBS_QUAL	mandatory	One to three letters
OBS_REV	optional	Boolean
OBS_CONF	optional	One letter
OBS_TYPE	mandatory	One or two letters
DATE_JO	optional	YYYY-MM-DD

Observations value

The observations are shown in anti-chronological order, the most recent ones listed first.

A dot is used as decimal mark, and there is no thousands separator: OBS_VALUE="1346.4" for example.

When the value is missing, OBS_VALUE="NaN" (« Not a Number ») is displayed.

Time period (marked as TIME_PERIOD) has a different format depending on the series frequency (YYYY stands for the year and n the period number) :

frequency	format	example	meaning
Annual	YYYY	2013	year 2013
Quarterly	YYYY-Qn	2014-Q3	third quarter of 2014
Monthly	YYYY-nn	1990-09	September 1990
Bi-monthly	YYYY-Bn	2010-B4	from July 1, 2010 to August 31, 2010
Semi-annual	YYYY-Sn	2012-S2	from July 1, 2012 to December 31, 2012

URL syntax

The HTTPS request for accessing the data follows the pattern defined by the [SDMX 2.1 specifications, section 7](#), part 4 « Data and Metadata Queries ».

The URL follows thus the pattern

<https://bdm.insee.fr/series/sdmx/data/dataflow/key?optionalparameters>

The path segment *dataflow* is used to identify the requested dataflow. In this case, the agency always being FR1 and the version 1.0, only the alphanumeric identifier can be provided. Therefore, the two following examples are equivalent: « FR1,IPI-2010-,1.0 » or just « IPI-2010 ».

The path segment *key* is optional and can be used to select one or several among all available series for the dataflow. It is built according to the dimensions defined in the DSD connected to the group. For each dimension, the requested values are indicated in this key and separated with a « + » character if several are chosen. The dimensions are separated with a « . » character and must respect the order specified in the DSD. If a dimension is left blank, it will not be used to filter the series and all possible values will be provided.

For example, for the IPI-2010 family (Industry Production price indices with 9 dimensions) (6 mandatory with NAF2, AUTRES_REGROUPEMENTS, BASIND, The key could be as follow :

- [/data/IPI-2010/...20-4.....](https://bdm.insee.fr/series/sdmx/data/IPI-2010/...20-4.....): retrieve all the series for the nomenclature item 20,4 from the NAF (nomenclature of french activities), i.e. 4 series including 2 annual and 2 monthly by which one seasonal and working days adjusted
<https://bdm.insee.fr/series/sdmx/data/IPI-2010/...20-4.....>
- [/data/IPI-2010/M...20-4.....](https://bdm.insee.fr/series/sdmx/data/IPI-2010/M...20-4.....): retrieve the two monthly series from NAF 20.4
<https://bdm.insee.fr/series/sdmx/data/IPI-2010/M...20-4.....>
- [/data/IPI-2010/...20-4....CVS-CJO.](https://bdm.insee.fr/series/sdmx/data/IPI-2010/...20-4....CVS-CJO.): retrieve the SA-WDA series for Naf 20.4 i.e. series 001654579 (idbank)
<https://bdm.insee.fr/series/sdmx/data/IPI-2010/...20-4....CVS-CJO.> (mind the last dot)

If this *key* part is missing from the URL or if it is equal to « all », every series of the group is returned.

The *optional parameters* are used to limit the amount of data sent back for each series. If they are not specified, all available values are displayed. The usable parameters are the following:

- **startPeriod**: beginning of the reporting period: 2001, 2010-Q3, 1990-09, ...
- **endPeriod**: end of the reporting period: 2014, 2013-S2, ...
- **firstNObservations**: integer n allowing to get the n oldest values of the series
- **lastNObservations**: integer n allowing to get the n most recent values of the series

- **includeHistory**: boolean to access the previous releases
- **updatedAfter**: starting point for querying the previous releases

The possible combinations of these parameters are:

- startPeriod + endPeriod: all observations included in the two periods are expected
- startPeriod + firstNObservations: the n oldest values starting from startPeriod
- endPeriod + lastNObservations: the n most recent values until endPeriod
- includeHistory + updatedAfter: to retrieve all the releases starting from updatedAfter

Last, the parameter named **detail** can be used to get only a part of the dataset, among its four different components: dimensions identifying the series, attributes of the series, attributes of the observations and observations value. The possibilities are listed below, depending on the value taken by the parameter:

detail	elements in the response
full	series dimensions + series attributes + observations attributes + observations value
dataonly	series dimensions + observations value
serieskeysonly	series dimensions
nodata	series dimensions + series attributes

If the parameter is not specified or is different from the 4 cases above, all possible elements are returned (same as *full*).

The client can select the output format by content negotiation, using the *Accept* header value in the HTTP request. For now, the possible values are (can be enriched later):

- application/vnd.sdmx.structurespecificdata+xml;version=2.1: StructureSpecificData format
- application/vnd.sdmx.genericdata+xml;version=2.1: GenericData format
- by default if not specified: StructureSpecificData format

Response structure

Two different output formats are offered for the BDM time series:

- StructureSpecificData: format linked to a specific DSD, the id of the various elements (dimensions, attributes) are directly used to define the xml attribute name

- GenericData: heavier format but with always the same structure for every DSD, the id of the various elements are only used as values of the xml attributes; in return, there are much more xml nodes

In the StructureSpecificData format, each series is represented in a <Series> element with as many xml attributes as there are sdmx dimensions and attributes for the series. It has n children <Obs> composed of the xml attributes TIME_PERIOD, OBS_VALUE, OBS_STATUS OBS_STATUS, OBS_TYPE, OBS_QUAL and potentially OBS_CONF, OBS_REV and DATE_JO.

In the Generic format, each series has 3 possible types of children:

- <SeriesKey>: an identification key with a value for each dimension
- <Attributes>: an optional list of attributes for the series
- <Obs>: a set of observations, with a time dimension element <ObsDimension>, a value element <ObsValue>, an optional list of attributes <Attributes> with the same presentation as the series-level ones

The SDMX message includes a header (XML element <Header>) before the data providing general information about the service, structured in several elements:

- <ID>: unique id of the response
- <Test>: boolean, always false
- <Prepared>: response building date and time, as yyyy-mm-ddThh:mm:ss
- <Sender>: FR1, identifier of Insee
- <Structure>: reference to the DSD used, with its 3 components agency, id and version
- <Source>: Insee.fr

IV. Access to structural metadata

Several resources, useful for understanding the meaning of the time series and its data, are available: Dataflow, Data Structure Definition, Concept Scheme, Codelist, Category Scheme.

URL syntax

The HTTP request for accessing the structural metadata follows the pattern defined by the [SDMX 2.1 specifications, section 7](#), part 4 « Structural Metadata Queries ». The URL follows the pattern:

<https://bdm.insee.fr/series/sdmx/resource/agency/identifier/version?optionalParameters>

The path segment *resource* can take the values: dataflow, datastructure, conceptscheme, codelist or categoryscheme.

The *agency* will always be FR1, the *identifier* is specific to each resource and the *version* can only be 1.0.

If the keyword *all* is used as the identifier, all the objects of the requested type are returned.

Some path segments can be excluded from the URL:

- <https://bdm.insee.fr/series/sdmx/resource/FR1/identifier>: equivalent to version=1.0
- <https://bdm.insee.fr/series/sdmx/resource/FR1>: equivalent to version=1.0 and identifier=all
- <https://bdm.insee.fr/series/sdmx/resource>: equivalent to version=1.0, identifier=all and agency=FR1

With the *optional parameter* references the client can get the objects linked to the requested resource. The possible values are parents, parentsandsiblings, children, descendants, all or directly the name of a specific resource like codelist for instance.

Response structure

The service response will be in SDMX Structure format, with the MIME type *application/vnd.sdmx.structure+xml*.

The root node is <Structure>, it contains two different childnodes <Header> and <Structures>. The requested resources are listed in <Structures> with a different element for each possible type (<Dataflow>, <DataStructure>, <ConceptScheme>, <Codelist>, <CategoryScheme>).

The header provides the following information:

- <ID>: unique id of the response
- <Test>: boolean, always false
- <Prepared>: response building date and time, as yyyy-mm-ddThh:mm:ss
- <Sender>: FR1, identifier of Insee
- <Source>: Banque de données macro-économiques

Examples

The request <https://bdm.insee.fr/series/sdmx/dataflow/> is used to retrieve all available dataflows, with information to make easier the data retrieval: number of series included in the dataflow, link to the equivalent group of series in the BDM website.

The <Dataflow> elements mention the DSD used to organise the corresponding datasets. So, with the output of the first request, the user can get the DSD linked to the dataflow IPC-2015-COICOP. It is also named "IPC-2015-COICOP" and can be retrieved at: <https://bdm.insee.fr/series/sdmx/datastructure/FR1/IPC-2015-COICOP>.

To understand the DSD, the associated concepts and codelists are required.

The concept scheme is available at: <https://bdm.insee.fr/series/sdmx/conceptscheme/>

The codelist CL_NATURE can be found at:

https://bdm.insee.fr/series/sdmx/codelist/FR1/CL_NATURE

V. Access to previous releases

For a limited set of series available through the webservice, an access to their previous releases is provided. At any change in the dataset (adding or deleting one value, changing values), a new version is released. The old version is stored and remains available by its original dissemination date.

To visualize the different datasets for a series, two parameters need to be used in the optionalParameters at the end of the URL. These two parameters are "includeHistory" and "updatedAfter".

The IncludeHistory Parameter :

The parameter includeHistory=true provides all the data releases.

Example: https://bdm.insee.fr/series/sdmx/data/SERIES_BDM/001710943?includeHistory=true

If the parameter value is not "true", the parameter is ignored and, in this case, the webservice retrieves only the latest release.

If the series is not concerned by revisions, the webservice response is:

"Aucun résultat ne correspond à cette requête"
(no data match this request).

The UpdatedAfter Parameter:

It is used to choose the depth of the history

Adding the updatedAfter parameter provides all the releases after the specified date.

The date is formatted as YYYY-MM-DDThh:mm:ss where YYYY stands for the 4-digit year, MM for the 2-digit month, DD for the 2-digit day, hh for the 2-digit hour, mm for the 2-digit minute, ss for the 2-digit second.

Example : <https://bdm.insee.fr/series/sdmx/data/IPC-2015/M.....ENSEMBLE...CVS.2015?includeHistory=true&updatedAfter=2017-07-11T08:45:00>

By adding this UpdatedAfter Parameter, the webservice response shows two other attributes in the XML tag `<message:DataSet>` :

1- validFromDate = to indicate the beginning of the time period from which the dataset is disseminated the dissemination time for the dataset;

2- action = this attribute can take one of these 3 values :

Append : when some new points were added but non revised

Replace : when the release contains some revised figures

Deleted : when some points were deleted

NB: the action attribute for the last release appears in the header part, `<message :DataSetAction>` tag:

`<message:DataSetAction>Append</message:DataSetAction>` (when one or more points were added without changing any older point)

Tracking updates

a HEAD request can also be made by using `If-Modified-Since=AAAA-MM-JJThh:mm:ss`.

If the dataset was modified, the webservice retrieves the data, with a 200 HTTP status. If not, the webservice response is a 304 status with an empty HTML page.

Release calendar

The daily updates are listed in one RSS 2.0 feed.

It is limited to the updates occurred in the last 30 days. The upcoming releases are not mentioned.

The RSSfeed provides a response with a content type *application/rss+xml* and is available at the following URL:

<https://bdm.insee.fr/series/sdmx/rss/donnees>

This calendar operates in pulling mode only. The client will have to regularly request it to be informed of the last updates.

The output for the data updates consists of a list of elements with:

- a title reminding the id and the name of the updated dataflow
- a reference to this dataflow
- a publication date which corresponds to the last update date of the time series

The output for the structural metadata updates offers several categories:

- codelist update (change of an existing item or addition of a new code)
- update or addition of a concept in the conceptscheme
- update or addition of a dataflow

Error handling

The error messages are compliant with the specifications [SDMX 2.1, section 7](#), part 5 « Standard Errors for SDMX Web Services ».

The response is an XML document with a technical code and a message, the HTTP status code is greater than 400. The possible cases are the following:

SDMX code	reason	status code
100	No results found	404
140	Syntax error	400
500	Internal server error	500
	Response size exceeds service limit, the amount of data requested must be limited	
510		413

For example in case of wrong identifier, the response message is:

<https://bdm.insee.fr/series/sdmx/datastructure/FR1/wrong-id/>