



# CIM Internet study

Methodology

Edition April 2020



CIM – Centre d’Information sur les Médias  
Avenue Herrmann-Debrouxlaan 46 - 1160 Bruxelles  
Tél. : 32 2 661 31 50 - Fax: 32 2 661 31 69  
E-mail : [internet@cim.be](mailto:internet@cim.be)  
URL : [www.cim.be](http://www.cim.be)



## Content

<b>PREFACE</b> .....	<b>5</b>
<b>SUMMARY</b> .....	<b>6</b>
<b>INTRODUCTION</b> .....	<b>7</b>
<b>STAKEHOLDERS</b> .....	<b>8</b>
CIM INTERNET TECHNICAL COMMITTEE .....	8
THE RESEARCH INSTITUTE.....	8
SUBSCRIBERS .....	8
RULES .....	8
<b>THE CIM INTERNET STUDY</b> .....	<b>9</b>
<b>COLLECTING POPULATION DATA</b> .....	<b>9</b>
<b>COLLECTING CENSUS TRAFFIC DATA</b> .....	<b>10</b>
2.1. SCRIPTING .....	10
<b>TRAFFIC PUBLICATION</b> .....	<b>14</b>
3.1. METRICS.....	14
3.2. PUBLIC RESULTS .....	14
3.3. GEMIUSPRISM.....	15
3.4. EGEMIUS .....	16
3.5. OFFICIAL CIM REPORTS .....	16
<b>COLLECTING AUDIENCE DATA</b> .....	<b>17</b>
4.1 RECRUITMENT.....	17
4.2 VALIDATION RULES.....	18
4.3 PANEL SIZE .....	19
<b>AUDIENCE MODELLING</b> .....	<b>20</b>
5.1. METHODOLOGICAL CHALLENGES.....	20
5.2. CALCULATING THE DAILY INTERNET POPULATION .....	21
5.3. CALCULATING DAILY AUDIENCE RESULTS .....	22
<b>AUDIENCE PUBLICATION</b> .....	<b>28</b>
6.1. PUBLICATION CONDITIONS.....	28
6.2. ACCESS TO THE RESULTS.....	28
<b>CONTROLLING THE CIM INTERNET STUDY</b> .....	<b>29</b>
7.1. CHECKING TRAFFIC DATA .....	29



7.2. CHECKING AUDIENCE DATA .....	29
<b>ANNEXES.....</b>	<b>30</b>
<b>ANNEX 1. INTAKE SURVEYS. ....</b>	<b>30</b>
CIM INTERNET PC COOKIE PANEL QUESTIONNAIRE FOR ADULTS (18+) .....	30
CIM INTERNET TABLET AND SMARTPHONE PANEL QUESTIONNAIRE FOR ADULTS (18+) .....	36
CIM INTERNET PC COOKIE PANEL SIMPLIFIED QUESTIONNAIRE FOR 12-17 YEAR OLDS.....	36
<b>ANNEX 2: PUBLISHED SOCIO-DEMOGRAPHIC VARIABLES.....</b>	<b>37</b>
<b>ANNEX 3: CALCULATION OF SOCIAL GROUPS .....</b>	<b>40</b>

## PREFACE

---

Avant toute chose, je tiens à remercier celles et ceux qui ont la curiosité et le courage d'aborder la lecture de ce précieux document. Même si la tâche est parfois ardue, lire la méthodologie d'une étude permet de mieux comprendre ses résultats et de s'assurer qu'ils sont bien représentatifs de la réalité. Ces deux points sont essentiels car la mesure des médias digitaux souffre encore trop souvent de confusion, d'erreurs d'interprétations ou de mauvaise réputation.

Les pages qui suivent vont donc faire la lumière sur l'étude CIM Internet, sous toutes ses coutures : la partie « census » qui mesure le volume de trafic sur sites « CIMés », la partie « audience » qui modélise le profil des visiteurs de ces sites et permet de calculer des performances sur des cibles socio-démographiques, l'accès au résultats (publication publique, rapports réservés aux souscripteurs de l'étude, outils d'analyse ou de media-planning disponibles).

Ce document peut également servir de référence pour les nouveaux venus qui souhaitent participer à cette étude. On y retrouve des liens vers des documentations techniques destinées aux webmasters, les règles à respecter, les définitions des principales mesures...

*The pace of change has never been this fast, yet it will never be this slow again.*

Cette édition 2020 de la méthodologie de l'étude CIM Internet, tient compte de toutes les modifications jusqu'au 1<sup>er</sup> avril 2020 :

- l'amélioration constante de la mesure de la vidéo avec la publication quotidienne des résultats d'audience (nécessaire pour compléter les audiences TV par les audiences de l'Online Vidéo),
- la prise en compte des nouveaux comportements concernant l'acceptation ou le refus des cookies,
- l'amélioration de la mesure des applications,
- l'amélioration du calcul du nombre de cookies et du nombre de surfeurs,
- la possibilité d'évaluer et d'optimiser les plans de campagnes sur des périodes sur-mesure (versus période d'un mois obligatoire),
- la nouvelle interface de publication des résultats...

A court terme, le calcul de performances de la publicité digitale et les mesures d'audience cross media figurent parmi nos objectifs.

A moins court terme, la mesure du digital continuera de s'adapter aux évolutions constantes des browsers (vers un monde sans cookie), des médias (fusion des plateformes TV et digitales, développement de l'audio digital...), du comportement des belges (nouvelles façons de consommer les médias, adoption de nouveaux objets connectés...), de la technologie (data, intelligence artificielle, machine learning), de la législation... L'agilité sera essentielle, car il est impossible de prévoir quelles nouvelles inventions apparaîtront dans les prochaines années. En revanche, ce que nous savons, c'est que le besoin de construire des références et d'établir des currencies communes aux différents acteurs du marché (digital ou non) restera essentiel. C'est ce qui permettra de (re)construire la confiance des annonceurs envers les médias digitaux et de renvoyer les cowboys dans les cordes. Une mesure fiable et neutre est essentielle au développement d'un marché sain.

Enfin, tout ceci serait impossible sans l'implication du CIM, de sa structure permanente, et des membres de la Commission Technique Internet. Nous pouvons tous remercier ces professionnels qui contribuent avec motivation à l'amélioration constante de la mesure des médias en Belgique.

Stéphanie Radochitzki,  
President of the CIM Internet Technical Committee

## SUMMARY

---

Since 2014, the CIM works with Gemius for the CIM Internet study. This study measures the traffic and audiences of websites, players and apps.

### Traffic

The CIM Internet traffic measurement is based on a 'site centric' approach, which requires media subscribers to the study to install a Gemius script on all pages of the sites, on their apps or their players that they want to be measured. When a subscriber has implemented the script, the institute (Gemius) will receive a hit each time a surfer visits a page or performs an action in a player or app. Gemius measures these page requests or actions and reports also visits, time and number of unique browsers. Traffic can be tracked on PC, laptop, smartphone and tablet.

Traffic data are published on a daily basis : Page/Stream Requests and visits (sessions) for sites, apps and players are available to the public, the details are only available for the CIM subscribers.

### Audience

The traffic measurement does not reveal who is behind a page request. To answer this question Gemius created a cookie panel : surfers on the participating websites receive an invitation to complete the CIM Internet study survey with questions on their socio-demographics and internet use. At the end of the survey a cookie is placed on the surfer's device and he becomes a panelist.

We distinguish 3 different panels depending on the device the panelist used to fill in the survey : PC (desktop + laptop), tablet and smartphone. There is a small overlap between those panels.

Within the PC panel, a distinction is made between the PC Work and the PC Home panel. Gemius typically measures only the home PC or only the work PC (but rarely both PC's). This means that for panelists that use a PC both at home and at work only 'half' of their internet activity is measured. To fill in the missing half, work PC data from another panelist are ascribed to a PC home panelist, or vice versa, home PC data from another panelist are ascribed to a PC work panelist. This newly made gross PC panel is then fused with the tablet and the smartphone panel, respecting duplications in device use known from the CIM Establishment Survey, to form the total cookie panel (+/- 80.000).

Finally, the total panel is weighted both on socio-demographics, as derived from the Golden Standard and Virtual Population (more information on [www.cim.be/nl/golden-standard](http://www.cim.be/nl/golden-standard) and [www.cim.be/fr/golden-standard](http://www.cim.be/fr/golden-standard)), and on surfing behavior, as derived from the census traffic data.

The audience results are published on a daily basis and made available to subscribers only in certified softwares ([www.cim.be/nl/internet/contact-internet](http://www.cim.be/nl/internet/contact-internet) and [www.cim.be/fr/internet/contact-internet](http://www.cim.be/fr/internet/contact-internet)).

## INTRODUCTION

---

Founded in 1971 from the merger of OFADI (first authentication agency for the distribution of press titles in Belgium) and CEBS (first Belgian agency for audience measurement), the CIM is an association that aims to provide reference figures for the Belgian advertising market. The data collected by CIM are primarily intended for members who co-finance the studies. They come from three families: advertisers, intermediaries (mostly media agencies) and media. However, some results are also publicly available on the CIM website ([www.cim.be/nl/internet/resultaten-belgie](http://www.cim.be/nl/internet/resultaten-belgie) and [www.cim.be/fr/internet/resultats-belgique](http://www.cim.be/fr/internet/resultats-belgique)).

CIM publishes internet traffic data for Belgian and Luxembourgian websites since 2000. The publication of Belgian internet audiences started in 2005. Since June 2014, the research partner for the CIM Internet study is Gemius.

As the digital world constantly evolves, the scope of the CIM Internet study has also evolved over time. The CIM Internet study currently consists of the following elements on PC/laptop, tablet and smartphones (other devices, e.g. connected TV or gaming devices are currently filtered out):

- Traffic measurement for content, at census level (= total universe, not sample or panel based)
  - o for websites on PC/laptop, tablet and smartphone,
  - o for HTML5, native and hybrid apps on tablets and smartphone,
  - o for audio and video players on websites or in apps.
- Audience measurement for content, based on PC/laptop, tablet and smartphone cookie panels.
- Traffic and audience measurement for advertising are currently under development.

A basic principle of the CIM Internet study is that it measures only digital assets (sites, apps, players) that are subscribing to the study. The site centric approach chosen by CIM requires that all digital assets are scripted or tagged by the publisher. Evidently, the scripting may not harm the user experience: it cannot cause a visible delay and it is executed asynchronously to avoid blocking a page, app or player if something goes wrong in the measurement.

## STAKEHOLDERS

---

### CIM Internet Technical Committee

The Technical Committee Internet supervises the study. On 1/1/2020, it was composed as follows:

- Chairman : Stéphanie RADOCHITZKI (Mediabrand)
- Members :
- Adrien DE WALQUE (RTL Belgium)
  - Nicolas MAGAIN (Space)
  - Mélanie PIERLOT (Rossel)
  - Sébastien REPS (Maxus)
  - Bart SPIESSENS (Pebble Media)
  - Bernd STRAUVEN (DPG media)
  - Karel VERBRUGGE (Mediahuis)
  - Frédéric WATELET (RTBF)
- CIM staff :
- Stef PEETERS (General Manager)
  - Kurt BAVEGHEMS (Head of TV & Digital Research)
  - Julie BADISCO (Assistant Project Manager)
  - Arno DEPOORTERE (Assistant Project Manager)

The CIM staff takes care of the relations with the subscribers and, together with the research institute, they monitor the correct implementation of all technical requirements.

### The Research Institute

Since 1st of June 2014 Gemius is the research institute responsible for the data collection, analysis and modeling of results of the CIM Internet study. Gemius is specialized in internet measurement, and provides the currency study in several European countries (see [www.gemius.com](http://www.gemius.com)).

### Subscribers

Subscribing CIM-members co-finance the study. Only these companies have access to the full results. They should handle the data with care and use them only within their normal commercial activities. The list of subscribers is updated on the CIM website: [www.cim.be/nl/internet/lijest-van-intekenaars](http://www.cim.be/nl/internet/lijest-van-intekenaars) in Dutch, [www.cim.be/fr/internet/liste-des-souscripteurs](http://www.cim.be/fr/internet/liste-des-souscripteurs) in French.

Technical partners and software houses have no right to share data with third parties. Any company that wishes to subscribe to the CIM Internet study can obtain all necessary information on the website [www.cim.be](http://www.cim.be) or from the CIM staff ([internet@cim.be](mailto:internet@cim.be)).

### Rules

Subscribers have to follow a number of technical rules and publication conventions set by the Technical Committee and approved by the Board of CIM. The current version is available on the CIM site ([www.cim.be/nl/internet/reglement-internet](http://www.cim.be/nl/internet/reglement-internet) in Dutch, [www.cim.be/fr/internet/reglement-internet](http://www.cim.be/fr/internet/reglement-internet) in French).



## THE CIM INTERNET STUDY

---

### Collecting population data

Most media currency studies of CIM use a rigorously selected representative sample of respondents (e.g. Press, Cinema and Radio) or a carefully composed panel (e.g. TV). This is not the case in the CIM Internet audience study that uses a continuous online survey. Recruitment is based on low frequency pop-ups that appear randomly on participating websites and invite the surfers to share sociodemographic data. Surfers that respond have selected themselves and self-selected samples are rarely representative (see also 3. Collecting profile and device usage data).

To know the real structure of the Belgian internet population, CIM has established an independent, representative, yearly study that monitors trends in possession and use of media equipment, and trends in media consumption for the total Belgian population. In the case of internet, it goes as far as the general surfing frequency on Belgian sites or the usage frequency of online videos... This so called “Establishment Survey” consists of 6.000 interviews a year (80% face to face recruitment, 20% online panel recruitment).

The data of the Establishment Survey are used to enrich the Virtual Population (VP, more information on [www.cim.be/nl/golden-standard](http://www.cim.be/nl/golden-standard)). The VP of CIM is a huge database that represents >11 million residents in Belgium, not only according to their sociodemographic characteristics but also in terms of key variables on device usage and media behavior. It serves as a reference not only for the internet study, but also for the CIM TV and Radio studies.

VP data on usage of pc/laptop, tablet, smartphone and their duplications, as well as their socio-demographic profile, will allow later on in the data processing to combine home and work pc data for surfers that use both (see 5.3.3), to model the reach of devices but especially their duplication (see 5.3.4.) and to correct sociodemographic recruitment biases of the online survey (see 5.3.4.d.).



## Collecting census traffic data

### 2.1. Scripting

The CIM Internet traffic measurement is based on a 'site centric' approach that requires the collaboration of the site, app and player owner to install the appropriate scripts and tags. All subscribing websites implement a small JavaScript tag into the source code of their website. Each time a webpage is requested by a browser, a call is sent to one of the measurement servers of Gemius. These servers retain the number of page requests and identify the browser.

If a browser visits one of the participating sites for the first time, a third party cookie is installed in the browser on the device used to visit the site. A cookie is a small text file that contains a unique numerical code. In case of a successful installation, the unique code will be sent to Gemius every time the surfer visits one of the participating (scripted) websites from the same browser on the same device. The cookie therefore allows to link page requests and visits (a uninterrupted series of pages requests by one surfer on one site) to different participating sites.

This type of measurement is not perfect. Several browsers do not accept cookies, either by default, because the user changed his settings, or because the user deletes his cookies occasionally. In such cases, a site centric measurement can correctly identify a page request but possibly not visits or a unique browser since the unique identification of the browser is unavailable. As the use of third party cookies will gradually disappear, research institutes continuously to search for alternative solutions to identify browsers.

Gemius uses proprietary BrowserID technology to identify more browsers, on top of third party cookies. The BrowserID technology combines local storage identifiers and 3rd party cookies to identify returning browsers. All Page Requests generated by a given BrowserID on a given website are grouped into a logical set to reflect an uninterrupted Visit.

The next step in identifying returning browsers is the use of 1<sup>st</sup> party cookies. These are not under threat since the correct functioning of the site and the memory of user setting depend on it. Since January 2020, Gemius uses 1<sup>st</sup> party cookie information to identify returning browsers on a given site, even when no 3<sup>rd</sup> party cookie information is available.

This, however, does not solve the issue of cross-site (also called cross-domain) identification: identifying which surfers on site A also visits site B. To solve this issue, Gemius prepares a cookie matching project that should enable connections between different 1st party cookies and cross domain identification of browsers.

#### 2.1.1. Scripting of HTML pages

The Gemius script for HTML pages is a JavaScript tracking script with a unique identifier per site (or site section) and several additional fields that can contain additional page-specific information that connects to the Gemius servers whenever a page request is made. The script is asynchronous: the response time of the Gemius servers will not affect the loading of the webpage itself. The script supports mobile and regular webpages, in HTTP and HTTPS, as well as applications built in HTML5 code.

```

<script type="text/javascript">
<!--//--><![CDATA[//><!--
var pp_gemius_identifier = 'IDENTIFIER';

var pp_gemius_extraparameters = new Array('lan=EN', 'key=keyword',
'subs=subsection', 'free=free_field');

// lines below shouldn't be edited

(function(d,t)                                {try                                {var
gt=d.createElement(t),s=d.getElementsByTagName(t)[0],l='http'+((location.pr
otocol=='https:')?'s':'');
gt.setAttribute('async','async');gt.setAttribute('defer','defer');
gt.src=l+'://gabe.hit.gemius.pl/xgemius.js';
s.parentNode.insertBefore(gt,s);} catch (e) {}})(document,'script');

//--><![ ]>
</script>

```

Besides the unique identifier, 4 other parameters are available to describe the content of the webpage:

1. **Lan** = Language. Possible values: FR, NL, EN, GE, LU, OTHER.

Only the language parameter is mandatory.

**Key** = Keyword: value to describe the content of the page. Specification of the information in the section. A keyword is used:

- To include additional info on site structure that can't be derived from the section only.  
E.g.: Section sports – Keyword: Tour de France
- To recreate commercial packages.

The number of different keywords is limited to 200 per site or section.

2. **Subs** = Subsection: free field for subscribers who wish to add more detailed info on their website structure.
3. **Free** = a free field to be used for whatever information the subscriber wants to collect.

Data on language and keyword are reported to the market in all Gemius tools.

Subsection and the free parameter are only available for internal analysis by a subscriber in gemiusPrism (see 1.2.6.).

Participating sites also need to implement the CIM logo and disclaimer.

More information on the measurement script and the correct tagging of a website is available for subscribers with the appropriate login on the CIM website:

French: [www.cim.be/fr/internet/documentation-technique-pour-les-souscripteurs](http://www.cim.be/fr/internet/documentation-technique-pour-les-souscripteurs)

Dutch: [www.cim.be/nl/internet/technische-documentatie-voor-intekenaars](http://www.cim.be/nl/internet/technische-documentatie-voor-intekenaars)

Each site that enters the study is controlled by Gemius and by CIM staff who checks if the tagging is implemented correctly according to the CIM Internet rules.



### 2.1.2. Scripting of HTML5 or AJAX pages

A special version of the script is available to measure websites using HTML5 or AJAX. Subscribers with such websites can ask the Permanent Structure for the documentation by sending an e-mail to [internet@cim.be](mailto:internet@cim.be).

### 2.1.3. Scripting of HTML5, hybrid and native apps

HTML5 apps are scripted with the regular page scripting technology (see 2.1.1).

Two SDKs (“Software Development Kit”) are available to tag native or hybrid apps: one for iOS native and hybrid apps, and one for Android native and hybrid apps. Subscribers can tag sections in their apps by using an extraparameter (‘sectionID’).

The appropriate documentation and a unique app identifier will be sent on simple request to any subscriber declaring to have an HTML5, native or hybrid app (send requests to [internet@cim.be](mailto:internet@cim.be)).

### 2.1.4. Scripting of streaming in sites

The CIM streaming traffic measurement is also ‘site centric’, meaning that the owner should install measurement codes in the player (and eventually in the streams to identify the content).

All subscribed players embedded in websites are required to implement a JavaScript tag into the source code of their website. Each time a player is started, paused, forwarded, rewind or stopped by the surfer, a call is sent to the Gemius servers. They will retain each player action and identify the browser.

The player tag uses the same 3rd party cookie as the web page tags. This allows to link all stream actions and visits coming from one unique browser (without details on which stream was played).

The tagging of audio and video players is done at the level of the player, not the website or the app it is played in. A player can therefore be used on different websites, even non-participating websites, and still be measured. When a player is used on a websites that do not belong to the subscriber, the measured volume will be reported without details under “MNC” (measured, not communicated).

Tagging a player also requires an identifier that is unique for each player. The same identifier can be used for a player on the subscribers’ own websites and on third party websites (MNC), and for both streaming in a browser and streaming in an application.

Since September 2019, Belgian broadcasters add a unique code to their video streams, similar to the codes they use in the classic TAM TV study. This allows to identify content that has been shown on both on linear TV and online.

The streaming documentation is available for JavaScript controlled players on the CIM site:

French: <http://www.cim.be/fr/internet/documentation-technique-pour-les-souscripteurs>

Dutch: [www.cim.be/nl/internet/technische-documentatie-voor-intekenaars](http://www.cim.be/nl/internet/technische-documentatie-voor-intekenaars)

```

1 <script type="text/javascript">
2 <!--//--><![CDATA[//><!--
3 function gemius_player_pending(obj,fun) {obj[fun] = obj[fun] || function() {var x =
4 window['gemius_player_data'] = window['gemius_player_data'] || [];
5 x[x.length]=[this,fun,arguments];}}};
6 gemius_player_pending(window,"GemiusPlayer");
7 gemius_player_pending(GemiusPlayer.prototype,"newProgram");
8 gemius_player_pending(GemiusPlayer.prototype,"newAd");
9 gemius_player_pending(GemiusPlayer.prototype,"adEvent");
10 gemius_player_pending(GemiusPlayer.prototype,"programEvent");
11 gemius_player_pending(GemiusPlayer.prototype,"setVideoObject");
12 (function(d,t) {try {var gt=d.createElement(t),s=d.getElementsByTagName(t)[0],
13 l='http'+((location.protocol=='https:')?'s':''); gt.setAttribute('async','async');
14 gt.setAttribute('defer','defer'); gt.src=l+'://gabe.hit.gemius.pl/gplayer.js';
15 s.parentNode.insertBefore(gt,s);} catch (e) {}})(document,'script');
16 //--><![ ]>
17 </script>

```

### 2.1.5. Scripting of streaming in application

Since August 2016, the installation of stream tagging within an application can be done by using a Software Development Kit (SDK). The documentation for Android native and hybrid apps, or for iOS native and hybrid apps is available upon simple request to the Permanent Structure of CIM ([internet@cim.be](mailto:internet@cim.be)).

No separate identifier is needed for streams in apps. The identifier of the player can be used. Gemius has developed a solution whereby UserAgent technology detects if a stream is coming from an app or a browser. This does require, however, that the UserAgent for the applications is correctly set.

## Traffic publication

Traffic results of the site centric measurement are published in different reports and software tools.

### 3.1. Metrics

The CIM Internet traffic measurement offers a broad range of results. Software tools have a manual that explain the available metrics. A list of metric definitions is also available in the rules of the CIM Internet study. The following short overview offers the most commonly used metrics.

- Page request  
A file, or combination of files, sent to a unique browser after a request of this unique browser was received through the server of the site.  
Warning: Gemius uses the term 'Page Views' in its reporting tools instead of 'Page Requests'.
- Stream impression  
Each time that an audio or video stream is started, either by a user action or by autoplay.
- Visit (or Session)  
A consecutive series of Page Requests generated within one browser on a given domain (site). The time between two consecutive page requests cannot exceed thirty minutes.
- UB (Unique Browser): a browser identified on the basis of a unique cookie or BrowserID (BID).
- Real Users: an estimate of the number of people that are represented by the UBs (this estimate is based on the goodBID algorithm, see 5.2.2).

More internet terms or expressions can be found on the CIM website in the glossary:

- French: <http://www.cim.be/fr/glossary>,
- Dutch: <http://www.cim.be/nl/glossary>

and in the Internet Rules

- French: <http://www.cim.be/fr/internet/reglement-internet>
- Dutch: <http://www.cim.be/nl/internet/reglement-internet>

### 3.2. Public results

The CIM publishes limited traffic results on its website [www.cim.be](http://www.cim.be):

- For Belgian websites, apps and players Page/Stream impressions and visits (sessions) are available for Belgian and global traffic. These results are provided in the online tool eGemius (<https://e.gemius.com>), which is placed in an I-frame on the CIM site.
- For Luxembourg websites and apps, daily, weekly and monthly reports on Page Requests, sessions and Unique Browsers are available on [www.cim.be/fr/internet/resultats-luxembourg](http://www.cim.be/fr/internet/resultats-luxembourg).

The technical information (on browsers, devices, OS, etc.) are aggregated and displayed on [ranking.gemius.com](http://ranking.gemius.com). Belgian results are available via [be.ranking.gemius.com](http://be.ranking.gemius.com) and Luxembourg results are available via [lu.ranking.gemius.com](http://lu.ranking.gemius.com).

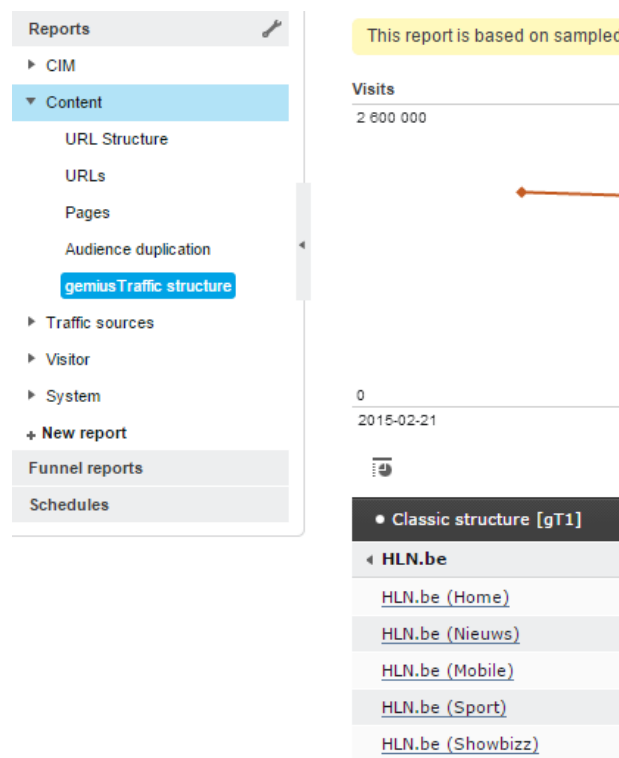
### 3.3. gemiusPrism

gemiusPrism is an online web analytics tool for subscribers that gives access to raw traffic data for proprietary sites, apps and players only (not to raw data of competitors). Only traffic created by bots is filtered out.

The tool reports near-live traffic data for participating sites, apps and players (maximum delay of 2 hours), allowing publishers to review their tagging, compare with internal data, or improve their website, players or apps. Prism data are no CIM currency data. Prism results are for internal usage only: no external communication of Prism data is allowed.

Traffic by section (if defined) can be analyzed

- either with a standard report “Content > gemiusTraffic structure”,
- or by selecting a section in any other report available in gemiusPrism.



An additional set of metrics is available for streaming projects, e.g. Stream views, Streaming time, Streaming time / visitor, Stream views / visitor, Streaming visitors... Within stream views there is a distinction available between net, gross and spot views in order to measure ad rolls.

Users can find a general manual on the usage of the gemiusPrism tool, and a guide tailored to the Belgian market, on the subscriber section of the CIM website:

French: <http://www.cim.be/fr/internet/documentation-technique-pour-les-souscripteurs>

Dutch: <http://www.cim.be/nl/internet/technische-documentatie-voor-intekenaars>.



### 3.4. eGemius

Starting the 2<sup>nd</sup> of April 2020, eGemius ([e.gemius.com](http://e.gemius.com)) is the webtool that offers both traffic and audience currency data for all published sites, apps, sections and players. From this date on, audience data will be based on the Constant Panel methodology that allows to calculate reach accumulation for any random selected time period within 3 months.

### 3.5. Official CIM reports

Official CIM reports are available on the CIM site for subscribers only.

Luxembourgian subscribers have access to traffic reports on a daily, weekly and monthly basis. They include detailed results onto the second level of subsection.

For the Belgian market, a monthly global audience report is available for backup purposes only.



## Collecting audience data

The CIM traffic measurement counts page requests within browsers, actions in apps, views in video players within the total surfer population... Unfortunately, this doesn't tell anything about gender, age or other socio-demographic characteristics of the surfer behind the browser, app or player. Yet, these audience data are essential in media planning.

This is precisely the reason why many internet audience studies, including CIM Internet, continuously collect profile data. Surfers are invited by low frequency pop-ups, that appear randomly on participating websites, to answer a limited number of questions on their devices, their internet usage and profile. The response rate is low but there a lot of surfers out there. So the problem is not so much the limited number of respondents but rather the unrepresentative self-selection.

As a result, the CIM Internet audience study is fundamentally different from other audience currency studies. The TV TAM study is based on a small but strictly controlled panel. The Radio, Press and Cinema studies are based on a larger, fairly controlled sample of individuals. In these methods, a 'simple' extrapolation to the general population (the universe) is sufficient to calculate profiled audiences.

The CIM Internet study uses another approach. By inviting surfers to profile themselves, we have profiled traffic from the browser and device the surfers used to fill in the pop-up questionnaire: as long as the identifier (cookie or BID...) used while responding to the questionnaire is kept, we collect surfing behavior from a profiled panelist on one device. In fact, Gemius has built over time three panels:

- a PC/laptop panel (split up into a 'home' and a 'work' section),
- a tablet panel,
- a smartphone panel.

The questionnaire used on mobile devices is slightly different than the questionnaire on pc's. Young people between 12 and 17 years receive a shortened version. All questionnaires can be consulted in the annexes of this document.

The advantage of this approach is two-fold:

- There is an observed (not a declared) relationship between socio-demographic data and browsing behavior.
- A panelist answers to the questionnaire responding to a popup on site A, but his profiles is applicable to all other sites he visits. This is an advantage for smaller sites that could never collect sufficient pop-up data on their own.

### 4.1 Recruitment

Surfers are invited to participate in the CIM Internet study through pop-up invitations shown on the participating websites. By responding to this request, they join the CIM Internet PC cookie panel, the CIM Internet Tablet cookie panel or the CIM Internet Smartphone cookie panel depending on the device they use to answer the questions.

The pop-ups follow a pseudorandomized pattern based on an algorithm that uses the following rules:

- Each day a subset of the Gemius cookie database is selected. This subset is maximum 45% of the entire cookie database.
- From this subset the invitation is only shown for cookies that meet several conditions:
  - The cookie or BrowserID must be at least 7 days old, to prevent surfers that delete their cookies regularly to become part of the panel.
  - If the cookie was already invited to participate without success, it will no longer be invited in the same timeslot (08:00 – 11:59, 12:00 – 14:59, 15:00 – 18:59, 19:00 – 21:59, 22:00 – 07:59).
  - In total, a given cookie cannot receive more than 2 invitations.
  - All subscribed websites participate. They don't control the odds of sending out an invitation.
  - The recruitment is continuous over all months of the year.
  - Only sites with children under 12 years as main target group can be excluded from the recruitment.
  - The CIM Internet Technical Committee can decide to temporarily raise the odds on specific websites that are selective on difficult to recruit profiles (e.g. to recruit young surfers on French websites).

There are three versions of the online questionnaire (see Annex 1):

- the standard PC/laptop questionnaire,
- a somewhat shortened version for Tablet and Smartphone,
- a version for 12-17 year old respondents without questions on main responsible for revenues, number and age of children, grandchildren...

Exceptionally, special recruitment actions are undertaken to recruit 12-17 or 18-24 year olds that are traditionally underrepresented (esp. on the French speaking side): a mailing to panel members with 12-17 year old children in the households, a short Facebook campaign.

## 4.2 Validation rules

In an online survey, inconsistencies should be easily prevented. E.g. if a person lives on his own, the question "main responsible for the income (MRI)" will not be asked but automatically set to 'Yes'. However, some (accidental?) inconsistencies remain possible. If not corrected, they will lead to the rejection of a respondent: e.g. a 17-year-old with a doctorate or a 106 year old participant.

As a matter of principle, responses of foreign residents and interviews completed on PCs in cybercafés are discarded. Foreign residents do not belong to the target group. Surveys from cybercafés can only be associated with session cookies and hence do not provide useful profile data.

There are additional validation rules: one set for the PC cookie panel, a shorter one for mobile panels as shown in the following table.

Validation rules for the PC/laptop panel and for the smartphone and tablet panels

	PC/Laptop	Tablet and smartphone
younger than 15 years old and no parental permission (STOP INTERVIEW)	X	X
age below 12 (STOP INTERVIEW)	X	X
age above 99	X	X
lives outside Belgium (STOP INTERVIEW)	X	X
wrong postal code	X	X
internet usage less often than once a month	X	X
fills in the questionnaire on device other than PC or laptop	X	
is not the main user of the PC/Laptop and doesn't have their own profile	X	
share profile on desktop computer	X	
share profile on portable computer	X	
fills in the questionnaire in a public place on desktop computer	X	
device other than tablet or smartphone		X
this tablet usage to visit internet websites less once a month		X
this smartphone usage to visit internet websites less once a month		X
respondent is not the main user of this tablet		X
respondent is not the main user of this smartphone		X
age below 15 and not a student	X	X
age below 20 and member of the general management	X	X
age below 30 and pensioner or retired	X	X
not working but declares filling in the questionnaire at work	X	
age below 18 and is a main income bringer	X	X
age below 20 and has a higher (bachelor education (or higher)	X	X
age below 22 and has an master education (or higher)	X	X

### 4.3 Panel size

The table below shows panel sizes from January 2019 to February 2020 both for PC and Mobile panels.

	PC/Laptop	Tablet/Smartphone	Total
2019.01	53.034	33.499	86.533
2019.02	51.676	32.784	84.460
2019.03	52.249	33.939	86.188
2019.04	51.650	34.651	86.301
2019.05	53.704	36.892	90.596
2019.06	50.660	36.064	86.724
2019.07	47.428	35.376	82.804
2019.08	47.144	34.793	81.937
2019.09	42.578	35.324	77.902
2019.10	42.752	36.116	78.868
2019.11	41.631	35.722	77.353
2019.12	40.685	34.473	75.158
2020.01	40.704	31.453	72.157
2020.02	39.941	31.327	71.268

## Audience modelling

### 5.1. Methodological challenges

Using a panel has the advantage that it creates long(er) term observations of surfing behavior across domains. Assuming that a solution is found for the different challenges discussed hereafter!

#### 5.1.1. The multi-cookie problem

Surfers who regularly delete their cookies are registered with several consecutive cookies. When in the PC Cookie Panel a sample of visitors is questioned, these surfers are recognized with only one cookie. Because of this incomplete time series, their profile is only connected to a part of the actual sites visited. This underestimation, and the fact that surfers who delete cookies might have a different profile, may result in a distortion of the socio-demographic profiles.

Gemius bypasses part of this problem by only showing the invitation pop-up to cookies or Browser id's that are at least 7 days old.

#### 5.1.2. The multi-browser problem

Surfers can use more than one browser on a certain device and multiple users can use the same browser on the same device. This is corrected by estimating the so-called J BRUS coefficient (see below 5.2.2).

#### 5.1.3. The multi-pc problem

Surfers can use the internet on more than one device, e.g. a desktop at work, and a laptop at home. This is also corrected by estimating the J BRUS coefficient (see below 5.2.2).

In 2015 there was no correction available for people surfing on a PC at work and on a (different) PC at home. Since 2016, a fusion of the home and work consumption for panel members declaring to surf both at home and at work/school, solved this problem (see below 5.3.3).

#### 5.1.4. The multi-user problem

When several people are using the same device, their internet traffic cannot always be attributed correctly. Therefore, surfers will only be included in any of the internet panels if they either:

- use their own login,
- use the same login as other people, but represent at least 50% of the total internet usage on that device.

People that do not meet this condition are screened out.

## 5.2. Calculating the daily internet population

The collected population data from the virtual population (see 1. Collecting population data) gives an estimate of the total number and profile of surfers on different devices for an average month. The total number of real people surfing on the internet, however, will vary from day to day. The first question is therefore whether variations in observed traffic, may serve to estimate the varying number of surfers per day?

Gemius refers to surfers as ‘Real Users’, more specifically defined as the number of internet users who visited at least one of the participating websites within the last 28 days. Real Users are published in eGemius and the planning softwares. In the Gemius’ methodology, the number of Real Users per day is estimated in several steps.

### 5.2.1. Real Users vs. Browsers

In all likelihood, the number of Real Users of a website will not be the same as the number of browsers used to visit that site. Cookie deletion and multi-device usage overestimate the number of Real Users, while device sharing underestimates it. Browsers are therefore no reliable estimate of Real Users. That’s why Gemius developed a dedicated algorithm to estimate Real Users.

### 5.2.2. The Real User algorithm

The 1st step in the algorithm is the evaluation of the number of **Real Browsers** that access the Internet.

- “good BrowserIDs” ( $BID_{good}$ ) number equals to number of browsers that were created before analyzed period and remained active till its end. The number of good browsers is estimated by summing the probability of each browser of becoming “good”. Such probability is estimated based on browser’s activity pattern and historical data for browsers with same activity pattern.
- For these good Browsers, the total number of pages viewed over all sites can simply be counted as sum of page views for each browser multiplied by its probability of becoming “good”. And we can therefore calculate the average total number of pages per good browser.
- If we divide the total number of page views (including the pages views from “bad” browsers for which the cookie was deleted at least once in the last 28 days) by the average number of pages for good browsers (b.), we have an estimate of the total number **Real Browsers** used yesterday.

The 2<sup>nd</sup> step in the algorithm is the estimate of number of **Real Users** behind the Real Browsers.

- To estimate the number of Real Users, we would need to know the average number of users per unique browser over all sites (the so called J-factor). This value can be derived from the number of Real Browsers for all sites combined, the number of surfers in the population as reported by the Virtual Population and the size of the node in question:  $J_{BRUS}(EC(n)) = (J_{str} - 1) \left( \frac{EC(n)}{EC(1)} \right) + 1$ , with  $J_{str} = \frac{Real\ Users_{Internet}}{EC_{Internet}}$ . For big nodes  $J_{node} \rightarrow J_{Internet}$ , for small nodes  $J_{node} \rightarrow 1$ .  $J_{Internet}$  is established for each device type independently.
- By multiplying the number of estimated browsers counted on a particular website by this J-factor, we get the number of **Real Users** for this particular website:  $RU_{Site} = EB_{Site} * J_{Site}$ .

A 3rd step in the algorithm is added from April 1st onwards : the **Facebook correction**.

- f. A growing number of pages is seen within the Facebook app that acts like a browser, but is not measured like a browser. The real number of browsers is therefore underestimated. By correcting this, the number of users per browser (J factor) is estimated more accurately for mobile devices.

A 4<sup>th</sup> step in the algorithm is also added from April 1<sup>st</sup> onwards : the PRES correction (PRES stands for Population and Reach Estimated Smoothly).

- g. This correction allows for **monthly variations in number of active surfers** based on daily variation in traffic data. This gives a more realistic estimate of the surfer population, including e.g. daily and seasonal variations, than fixing the total number of surfers according to structural population data provided by the Virtual Population.

### 5.3. Calculating daily audience results

#### 5.3.1. The Constant Panel

Starting the 1<sup>st</sup> of April 2020, audience results are produced daily for the Constant Panel. The Constant Panel is a virtual representation of the Belgian population with a limited size (+/- 120.000 records), not to be confound with the large Virtual Population of CIM (>11Mio records).

All records in the Constant Panel have the same weight. This weight is therefore the extrapolation factor used to calculate audiences for the total Belgian population. Each record receives socio-demographic and behavioural characteristics (like surfing or not; surfing the last 30d on Belgian sites...). Each day, the surfing behaviour observed in the pc, tablet and smartphone panels is ascribed to the virtual Constant Panel, respecting the duplication between devices and as much as possible the history of ascriptions (= any given virtual individual will receive as much as possible the observed behaviour of the same real panelist). Given the daily variation in internet population (see 5.2.), the status of individuals in the Constant Panel may vary between days from active to inactive surfers as in real life.

Before this daily ascription process is executed, however, a list of preliminary processing steps has to be performed.

#### 5.3.2. AB merging

Browser identification allows to relate Real Users from site A to Real Users from site B and count their duplication. However, such mechanism is not available for apps. Worse: since apps do not emit a pop-up questionnaire, their profile is unknown.

Gemius uses 'application-browser merging' (AB merging) to establish the relationship between sites and apps, and by doing so, create profile data for apps. The basic principle of AB merging is connecting activity of a panelist recruited and measured within a web browser with his activity in a measured application. This is done by pairing a Browser ID (representing the web browser used by the panelist) with an Advertising ID (identifying an application installed on the user's device), resulting in a new panel that contains activity on both web browsers and scripted apps. There are two steps in this

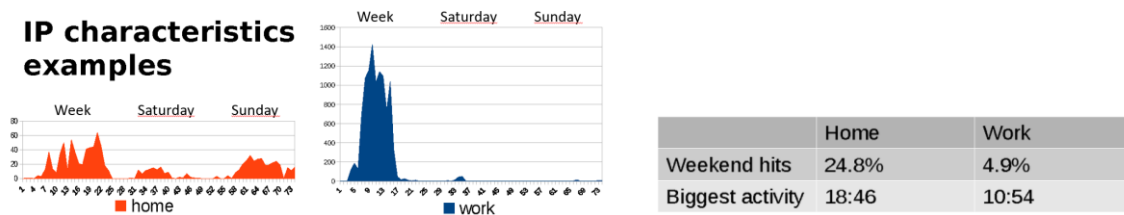
process: in first one, actual connections between browser and app on one device are detected (based on device characteristics (UserAgent), activity generated on matching IP address at matching times. In second step, app activity is filled for those users for which no true detection was detected (only for share of users allowing to cover applications reach on the market). Panelists for which the activity should be completed receives app activity of the user, who has www and app activity and whose www activity is most similar.

Since AB merging always takes the web browsing panelists as a starting point, app-only users can't be reached.

### 5.3.3. PC Home and PC Work ascription

In the PC/laptop panel, we have observations from home computers and from work computers, but (almost) no combined results from surfers that use both. Gemius therefore models the combined usage.

First, all PC page requests are classified as either PC Home page requests or PC Work page requests based on an IP classification model that takes the usage pattern of computers into account, e.g.



This model allows to:

- correctly classify 98% of the Home IPs to 'Home',
- correctly classify 83% of the Work/School IPs to 'Work'.

As there are more home IPs than work IPs, the model overall correctly classifies 94% of IPs. The Home-Work distribution by website will be used to calibrate the fusion of the PC panel data.

Secondly, PC Panel data will be split into PC Home and PC Work data based on the answers to the question *“How often have you used internet in the past month on each of the following places?”*

- *At Home*
- *At work*
- *At school or at university*
- *Mobile (on the street, underway or at public places)*
- *Elsewhere*

(1. Never, 2. Less than 3 days a week, 3. 4-6 days a week, 4. 7 days a week)

In a third step, home pc's and work pc's (as identified in the pop-up questionnaire) will be combined and weighted to represent subgroups within the PC surfer population as it is known from CIM's VP:

- Exclusively PC at home surfers (they will not be fused),
- Exclusively PC at work surfers (they will not be fused),
- PC Home panel members, declaring using both, that will be receivers of PC Work data,
- PC Work panel members, declaring using both, that will be donors for PC Home panelists,
- Surfers that are part both of the Home PC and the Work PC pane (the calibration panel).

#### 5.3.4. Fusing devices with BPS

Having separate surfing and profile data from PC/laptop, tablet and smartphone panel, does not allow to represent the behavior of surfers that use multiple devices.

- Luckily, respondents to the Gemius pop-up questionnaire indicate what type of devices they use to surf. So we know which part their online activity is missing.
- A small number of respondents fill in the questionnaire on several devices (spontaneously or after an invitation by mail), constituting a small calibration panel.
- Furthermore, we know the duplication between devices at the population level from CIM's Virtual Population.

This combination of “known incomplete” panel data with structural population data allows Gemius to fuse data from their 3 panels into a representative image of the population of single, dual, triple or more device users. This methodology is called the “Behavioral Panel Synthesis” (BPS) methodology.

The BPS algorithm consists in essence of four steps:

- a. Metrical Clustering
- b. Behavioral Distance
- c. Nearest Neighbor Merging
- d. Weighting

Each step is described shortly hereafter.

##### **a. Metrical Clustering**

First, each panel is clustered into subsets of similar users based on the socio-demographic variables gender, age and language and the devices they use. Only panelists from within one cluster will be merged. In other words, two significantly different panelists from different clusters will never be merged (e.g. a man with a woman, or a Dutch speaking with a French speaking). Panelists who use only one device, are excluded from the fusion process.

##### **b. Behavioral Distance**

There is a great variety of behavioral patterns (surfing) within a single cluster. To avoid merging significantly differing panelists, a probabilistic model is used whereby the distance between all the individuals is calculated. That distance is based upon their socio-demos as well as on their surfing behavior. This distance reflects the probability of being the same person for two panelists on different devices. This allows to merge only those panelists whose behavior indicates that they are likely to be the same person. In practice, this principle is build up as follows:

- To build such a model a Calibration Panel is required. It provides information about the internet usage of panelists across different devices. In February 2020 the calibration panel was composed of 468 panelists out of which 360 were enlisted in the PC and Smartphone panel, 70 in the PC and Tablet panel and 38 in the Smartphone and Tablet panel.
- Within the calibration panel, it is known for all panelists which websites he visited on all his devices within the analyzed period. This allows to establish for each calibration panelist positive connections between devices (websites visited by the same person on both devices) and false connections (websites visited on only one device).



- The pattern of connections (0 or 1) observed in the calibration panel, will be used to calculate the distance between calibration panelists that use different devices to see who are the best candidates for merging. This creates for each combination of devices a model of good and bad fits based on a logistical regression. The model is “self-training”. From month to month the algorithm builds upon the acquired information.
- Finally, the model will be applied to the appropriate combination of devices within a given cluster (e.g. Dutch Male 18-24). For each possible combination of panelists with one device with all panelists from another device a distance coefficient is calculated.

### **c. Nearest Neighbor Merging**

Within each cluster (e.g. Dutch Male 18-24) we merge panelists from different platforms as follows:

#### **1. PC \* Smartphone:**

- For any PC panelist, the probability is calculated of being the same person as each Smartphone panelist.
- The Smartphone panelist with the highest probability is merged with the given PC panelist and the Smartphone panelist was not already merged with too many PC panelists). The merged panelist keeps the sociodemos of the PC panelist.
- If there is no such Smartphone panelist, then the given PC panelist is not merged with any Smartphone panelist.

#### **2. PC \* Tablet:**

- Procedure 1 is repeated with Tablet panelists, instead of Smartphone panelists.
- The merged panelist keeps the sociodemos of the PC panelist.

#### **3. Smartphone \* Tablet:**

- Procedure 1 is repeated with Tablet and Smartphone panelists.
- The merged panelist keeps the Phone sociodemographics.

These three processes are set to produce the required number of device combinations according to CIM’s VP. This process results in a cross-device panel in which the full surfing activity on all devices is modelled for each panelist.

### **d. Weighting**

The last step of the BPS approach is the daily RIM weighting procedure that takes into account:

- The sociodemographic profile of users for each platform and for total internet.
- The device structure and profile, i.e. # of users and sociodemographic profile per device combination (e.g. PC only, PC + Mobile, PC + Smartphone...).
- The number of Real Users for each website on each platform.

### 1. Sociodemographic profile

Surfers in the constant panel are weighted to represent all surfers residing in Belgium, who were active during the last 30 days on Belgian sites, following this weighting matrix:

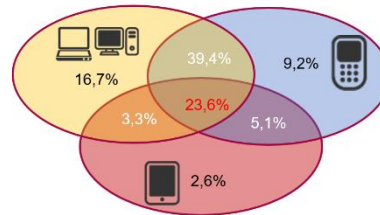
- Language (2)
- Gender(2) \* Language(2)
- Age(6) \* Language(2)
- Age(6) \* Gender(2)
- Education (6)
- Education (2) \* Language(2)
- Occupation (11)
- Activity(2) \* Language(2)
- Nielsen (8)
- Province(11) \* Gender(2)

Weighting objectives are updated twice a year based on CIM's VP (1<sup>st</sup> of January) or the latest Establishment Survey (1<sup>st</sup> of July).

### 2. Device structure and profile

The usage of different devices, and their duplication, is derived from CIM's VP. The table and graph below show the weighting objectives on the 1<sup>st</sup> of January 2020.

Internet population by device	
Total population	9.885.365
Total surfers last 30 days	7.905.829
PC	6.509.679
Tablet	2.714.173
Smartphone	6.064.073

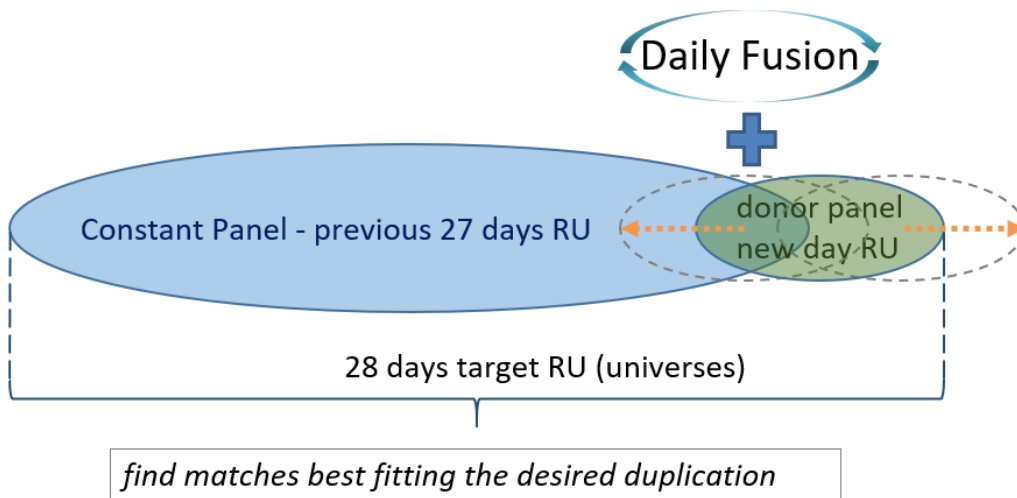


### 3. Real Users per website \* device.

The total modelling process should reproduce the expected number of Real Users by site \* device as established in 5.2.2. In other words, the sum of weights of all the panelists that visited a particular website should be equal to the expected number of Real Users.

### 5.3.5. Daily Fusion

Daily Fusion is a core step allowing to combine independent daily files with Constant Panel, where each virtual panelist has his own browsing history represented over time. The priority here is not to simply pair the most socio-demographically and behaviorally similar panelists (no seek for look-alikes). The goal is to keep socio-demographical clusters (to assure adequate demographical profiles), but inside of them, pair panelists in such a way, that it minimizes Real Users error compared to site-centric measurement. So now, Daily Fusion is responsible for long-term error reduction instead of weighting.



### 5.3.6. Calibration to census traffic data

The best modelling and weighting procedure will never produce traffic results that match perfectly the census traffic results. As it would be annoying to have slightly different traffic results according to the source and tool, the modelled traffic results are calibrated to match census traffic data at the level of page requests, visits and time.

## Audience publication

### 6.1. Publication conditions

In principle, each Belgian website, app or player that participates in the CIM Internet study and – in case of websites – that allows the correct display of pop-up questionnaires (see 4.), will automatically receive both traffic and audience data, as far as the app or player is tagged correctly.

Results from publications units for which the number of Real Users stay below 0,3% of the population are marked in grey.

### 6.2. Access to the results

Results from the CIM Internet audience study are only accessible to subscribers. They can be consulted in 3 different formats:

- Monthly global Excel report
- eGemius
- Media planning tools.

Certified softwares make the results available in planning software certified by CIM. The list of current certified software suppliers is available on the CIM website:

FR: [www.cim.be/fr/internet/fournisseurs-de-logiciel](http://www.cim.be/fr/internet/fournisseurs-de-logiciel)

NL: [www.cim.be/nl/internet/softwareleveranciers](http://www.cim.be/nl/internet/softwareleveranciers)

#### 6.2.1. Monthly Excel report

Each month, a global report with socio-demographical profiles for all publication units is published on the CIM website and available to subscribers of the study.

The report summarizes data on Belgian traffic with a socio-demographical split on gender, age, language, region, province, degree, employment status, profession, PRI, social group, number of people in the household, grandparent, age of (grand)children, PRP, internet devices, number of children in the household and urbanization.

#### 6.2.2. The eGemius reporting tool

eGemius is an online tool ([e.gemius.com](http://e.gemius.com)) released in April 2020 and accessible only to subscribers of the CIM Internet study. The tool contains both traffic and audience data for sites, sections, apps and players. Since the introduction of the Constant Panel (see 5.3.1), the user can calculate the reach accumulation for any freely selected period. Annex 2 shows the list of available socio demographics.

#### 6.2.3. Media planning tools

Since the introduction of the Constant Panel (see 5.3.1), software houses receive daily planning files. These files contain the same information as eGemius. Planning tools can produce crosstabs (as does eGemius) but also allow to plan and simulate digital advertising campaigns. The available metrics are Reach, Number of surfers, Views, Visits and Timespent (per member of a target group or for all members of a target group together).

## Controlling the CIM Internet study

Each aspect of the CIM Internet study is thoroughly tested by the research institute Gemius as well as by the CIM SPS and by the software house (GfK Probe).

### 7.1. Checking traffic data

The tagging of each new website/section/player/app is thoroughly tested before allowing it into the study. The tagging of a random selection of websites, players and apps is continuously checked, e.g.:

- Is the scripting correctly implemented?
- Is the scripting available on the entire website/app?
- Is the correct identifier being used?
- Is the mandatory extraparameter 'Language' present?
- Are the CIM logo and disclaimer available on the website/app?

Every day the CIM Internet staff checks results for Belgium in eGemius or gemiusPrism, results for Luxembourg in the offline reports. Checks are done on the availability and stability of the results. If a problem is discovered, the subscriber is contacted and asked to take the necessary measures.

### 7.2. Checking audience data

Before delivering the data to CIM, Gemius does a series of internal checks on the audience data:

- Checking for abnormal volumes of filtered out page requests (based on rules for allowed domains, autorefresh or use of iFrame),
- Monthly trend reports on validation rules at panel intake (e.g. how many respondents lost due to conflicting answers in the intake survey?)
- Monthly stability check on % of not-good BID
- Check size of 3 panels and compare with population trends (e.g. increase smartphone...)
- Check on the socio-demographical structure of the panel versus the data known CIM's VP
- Check on weighting efficiency and weights distribution (average, min., max. weight)
- Monthly trend reports for Total Internet (e.g. Estimated Cookies, Population, J coefficient)
- Monthly trend report on all websites (e.g. changes in PVs, RUs, panelists)

The Permanent Structure of CIM controls the following:

- Check for gaps in traffic data.
- Monthly overview ins and outs (site, section, app, player and networks) for software houses.
- Comparison with CIM's VP of the published results on weighted and unweighted variables for PC, Smartphone and Tablet audiences
- Check on stability of reach results for all websites x devices

On a monthly basis the audience data are controlled and reported to the CIM Internet technical committee. Reports are presented on the panel size, the recruitment rate, the socio-demographic composition of the panel, the efficiency of the weighting ...

## ANNEXES

### Annex 1. Intake surveys.

#### CIM Internet PC Cookie Panel questionnaire for adults (18+)

Age and place of living			
Q0	year (1900 TO current year - 12)	What is your year of birth?	----
Q11A		Do you live in Belgium?	1: Yes 2: No
		What province do you live in?	1: Antwerp 2: Brussels 3: Hainaut 4: Limburg 5: Liège 6: Luxembourg 7: Namur 8: East-Flanders 9: Flemish Brabant 10: Walloon Brabant 11: West-Flanders
Q11B	numeric	What is your zip code?	----
General internet usage			
Q1	single	How often do you usually use internet?	1: 7 days a week 2: 5 or 6 days a week 3: 3 or 4 days a week 4: 1 or 2 days a week 5: Less than 1 day a week 6: Less than 1 day a month
Q1b		How many hours do you use the Internet in a typical day?	1: up to one hour 2: 1-2 hours 3: 3-4 hours 4: 5-6 hours 5: 7 hours or more
Q1c	GRID (TABLE) QUESTION. One answer for each location.	How often do you generally use internet on each of the following places ?	A. Home B. At work C. At school or at university D. Mobile (on the street, underway or at public places) E. Elsewhere  1. 7 days a week 2. 4-6 days a week 3. <= 3 days a week 4. Never

Q2	multiple + open text	Over the past month, have you at any time used the internet via one of the following devices (e.g. to surf, send or receive e-mails, use online banking or social networks,...)?	<ol style="list-style-type: none"> <li>1. Desktop computer</li> <li>2. Portable computer (laptop)</li> <li>3. Tablet</li> <li>4. Smartphone (mobile telephone with internet access)</li> <li>5. Portable multimedia console (e.g. PlayStation Portable, Nintendo DS,...)</li> <li>6. TV player</li> <li>7. Another device: _____</li> </ol>
Q2b	If Q2 = 1	Over the past month, have you at any time used the internet on a <b>desktop computer</b> on the following places (e.g. to surf, send or receive e-mails, use online banking or social networks,...)? <i>Several answers possible</i>	<ol style="list-style-type: none"> <li>1. at home</li> <li>2. at work</li> <li>3. in school or at the university</li> <li>4. with friends or acquaintances</li> <li>5. somewhere else</li> </ol>
Q2c	If Q2 = 2	Over the past month, have you at any time used the internet on a <b>portable computer</b> on the following places (e.g. to surf, send or receive e-mails, use online banking or social networks,...)? <i>Several answers possible</i>	<ol style="list-style-type: none"> <li>1. at home</li> <li>2. at work</li> <li>3. in school or at the university</li> <li>4. with friends or acquaintances</li> <li>5. somewhere else</li> </ol>
Q3	single	What device are you using right now to fill out this survey?	<ol style="list-style-type: none"> <li>1. Desktop computer</li> <li>2. Portable computer (laptop)</li> <li>3. Tablet</li> <li>4. Smartphone (mobile telephone with internet access)</li> <li>5. Portable multimedia console (e.g. PlayStation Portable, Nintendo DS,...)</li> <li>6. TV player</li> </ol>
Q4	single	Are you the only person who uses the internet from this device?	<ol style="list-style-type: none"> <li>1: Yes</li> <li>2: No</li> </ol>
Q4b (If Q4=2)	single (both) Together on 1 screen	What is your share in the overall <u>internet</u> use from this device?	<ol style="list-style-type: none"> <li>1: I am the main user of this device</li> <li>2: I use this device to the same degree as other user(s)</li> <li>3: I make less use of this device than other user(s)</li> <li>4: I don't know</li> </ol>
Q5 (If Q4 =2)		Are different accounts (logins) being used on this device?	<ol style="list-style-type: none"> <li>1: Yes</li> <li>2: No</li> <li>3: I don't know</li> </ol>
If Q5=1		What account (login) do you use on this device?	<ol style="list-style-type: none"> <li>1: I use the same account (login) as other users</li> <li>2: I use my personal account (login)</li> <li>3: I don't know</li> </ol>
Q6 (If Q3=1 Vaste computer)	single if (Q6 = 3 OR Q6 = 4 OR Q6 = 5) STOP interview (screen out)	Where is this desktop computer located ?	<ol style="list-style-type: none"> <li>1. at home</li> <li>2. at work</li> <li>3. in school or at the university</li> <li>4. with friends or acquaintances</li> <li>5. somewhere else</li> </ol>

<b>Q7 (If Q3=2 laptop)</b>	single	Where do you mainly use this portable computer?	<ol style="list-style-type: none"> <li>1. at home</li> <li>2. at work</li> <li>3. in school or at the university</li> <li>4. on different locations</li> </ol>
		(Think about the internet browser by which you're filling in this survey on this device). Are you using also other browsers on this device? If so, how often?	<ol style="list-style-type: none"> <li>1: I only use this browser</li> <li>2: I use this browser more often than others</li> <li>3: I use this browser equally often to others</li> <li>4: I use this browser less often than others</li> <li>5: I don't know/hard to say</li> </ol>
<b>Q8 (If Q3=3 Tablet)</b>	single	Where do you mainly use this tablet?	<ol style="list-style-type: none"> <li>1. at home</li> <li>2. at work</li> <li>3. in school or at the university</li> <li>4. on different locations</li> </ol>



Basic socio-demographic variables			
Q9	single	What is your gender?	1. Male 2. Female
Q12	single	What is the highest level of education that you have reached successfully, either in day school or evening school?	1: primary school or no degree 2: lower general secondary education (first 3 yrs) 3: lower secondary technical, artistic or professional education (first 3 years) 4: higher general secondary education (last 3 yrs) 5: higher secondary technical, artistic education (last 3 years) 6: higher secondary professional education (last 3 years) 7: candidate, bachelor (academic or professional), graduate 8: university licentiate, master, post graduate, extra-university higher education (long type) 9: university license with met additional degree, master after master 10: doctorate with thesis
Q13	single	Which category responds the best to your current professional situation?	1: I am a pupil   student   in formation 2: I work full-time 3: I work part-time 4: I have temporarily suspended my professional occupation / I'm using up time credit 5: I have no employment at the moment (e.g. pensioned, jobless, ...)
Q14 (If Q13=5)	single	Which statement suits you best?	1: I am a houseman   housewife 2: I am incapacitated 3: I am jobless 4: I am in early retirement 5: I am pensioned 6: Other
Q15 (If Q13= 2 or 3 or 4)	single	What is your professional status (for your main occupation)?	1: independent 2: employed in the public sector 3: employed in the private sector
Q16 (If Q13= 2 or 3 or 4)	single	Which of the following categories corresponds the best with your profession?	1: farmer 2: artisan 3: merchant, industrialist 4: worker 5: clerk 6: middle management (e.g. head of department or division, ...) 7: member of the general management, senior executive (e.g. director, manager, ...) 8: free profession 9: freelance, self-employed
Q17 (IF Q16 = 1 or 2 or 3 or 6 or 7 or 8 or 9)	single	For how many employees are you responsible?	1: 0 2: from 1 to 5 3: from 6 to 10 4: 11 or more
Q18 (If Q16=4)	single	Are you a ...?	1: skilled worker 2: unskilled worker

<b>Q19 (If Q16=5)</b>	single	Are you a ...?	1: clerk with primarily office work 2: clerk with little or no office work
<b>Q20</b>	numeric 0-15	Can you indicate how many household members, EXCLUDING yourself, are living with you on a permanent basis or at least half of the time (e.g. during the week, one week out of two)	_ _  members
<b>Q21 (If Q20 &gt;0 for each)</b>	numeric 0-99	Can you indicate the age of each household member ?	_ _  year
<b>Q22 (If Q20 &gt;0)</b>	single	Are you usually the person in charge of the choice of brands for food, drinks and maintenance products in your household?	1: Yes 2: No
<b>Q23 (If Q20 &gt;0)</b>	single	Who is the main responsible for the income in your household (the person with the highest income)?	1: Myself 2: Someone else
<b>Q24 (If NOT Q23 PI is Myself - PI = MRI)</b>	single	What is the highest level of education that was reached successfully by the <u>main responsible for the income</u> , either in day school or evening school?	1: primary school or no degree 2: lower general secondary education (first 3 yrs) 3: lower secondary technical, artistic or professional education (first 3 years) 4: higher general secondary education (last 3 yrs) 5: higher secondary technical, artistic education (last 3 years) 6: higher secondary professional education (last 3 years) 7: candidate, bachelor (academic or professional), graduate 8: university licentiate, master, post graduate, extra-university higher education (long type) 9: university license with met additional degree, master after master 10: doctorate with thesis 11: I don't know
<b>Q25 (If NOT Q23 PI is Myself - PI = MRI)</b>	single	Which category responds the best to the current professional situation of the <u>main responsible for the income</u> ?	1: pupil   student   in formation 2: working full-time 3: working part-time 4: has temporarily suspended his/her professional occupation / is using up time credit 5: has no employment for the moment
<b>Q26 (If Q25 = 5)</b>	single	Which statement suits the <u>main responsible for the income</u> best?	1: houseman   housewife 2: incapacitated 3: jobless 4: in early retirement 5: pensioned 6: other

Q27 (if Q25 = 2 or 3 or 4)	single	Which of the following categories corresponds the best with the profession of the <u>main responsible for the income?</u>	1: farmer 2: artisan 3: merchant, industrialist 4: worker 5: clerk 6: middle management (e.g. head of department or division, ...) 7: member of the general management, senior executive (e.g. director, manager, ...) 8: free profession 9: freelance, self-employed
Q28 (if Q10 age >= 36 years)	single	Are you a grandfather or grandmother ?	1: Yes 2: No
Q29 (if Q28=1)	multiple	Do you have grandchildren in the age ranges below?	1: 0 to 2 year 2: 3 to 5 year 3: 6 to 11 year 4: 12 to 14 year 5: 15 to 17 year 6: 18 to 24 year 7: 25 year and older
Q30	single	What language do you usually speak at home?	1: Dutch 2: French 3: German 4: English 5: Arab 6: Spanish 7: Italian 8: Polish 9: Turkish 98: Other: ..... (specify)
Q31	single	Do you speak other languages at home?	1: Yes 2: No
Q32 (if Q31 = 1)	multiple	What other languages do you speak at home?	1: Dutch 2: French 3: German 4: English 5: Arab 6: Spanish 7: Italian 8: Polish 9: Turkish 98: Other: ..... (specify)
<b>Participation</b>			
<b>Q90</b>	single + open	Do you wish to participate in the sweepstake? (You can win ....)	1. Yes, if I win, please inform me via the following e-mail address [.....] 2. No, I do not wish to participate

## CIM Internet Tablet and Smartphone Panel questionnaire for adults (18+)

This version is a slightly shortened version of the PC version. Questions on the number of hours spent on the Internet on a typical day, accounts used on the device and the language(s) spoken at home are not asked in this version.

## CIM Internet PC Cookie Panel simplified questionnaire for 12-17 year olds

What is your year of birth?	
12-14 years: You have to ask your parents for permission if you want to take part in the study. Do you give permission for your child to take part in the survey?	1: Yes 2: No
Do you live in Belgium?	Yes/No
What province do you live in?	Antwerp Brussels Hainaut Limburg Liège Luxembourg Namur East-Flanders Vlaams-Brabant Waals-Brabant West-Flanders I do not live in Belgium
What is your zip code? (Only for Vlaams-Brabant)	
Generally where do you connect to the Internet <u>using desktop computer or laptop</u> ?	at home
	at work
	at school or at the university
	at friends or acquaintances
	other place
Over the past month, have you at any time <b>used the internet via one of the following devices</b> (e.g. to surf, send or receive e-mails, use online banking or social networks, ...)?	Desktop computer
	Portable computer (laptop)
	Tablet
	Smartphone (mobile telephone with internet access)
	Portable multimedia console (e.g. PlayStation Portable, Nintendo DS,...)
	TV player
Another device	
Who uses internet on the <u>device</u> you are on now?	1: mainly me 2: mainly other children 3. mainly adults
What is your gender?	Female/Male
Do you wish to participate in the sweepstake? (You can win <u>nice prizes</u> )	Yes, if I win, please inform me via the following e-mail address
	No, I do not wish to participate
Can CIM invite you in the future to participate in other media studies?	Yes, please invite me via the following e-mail address: _____
	No, I do not wish to be invited

## Annex 2: Published socio-demographic variables

Most socio-demographic variables are questioned directly (e.g. age, gender or professional activity of the person interviewed). Others are the result of a treatment on basic data (e.g. social groups, region). The following variables are available in eGemius and the planning files:

GENDER	MEN WOMEN
AGE IN YEARS	
LANGUAGE	FRENCH DUTCH
NIELSEN REGION	I   II   III   IV   V
PROVINCES	WALLOON BRABANT BRUSSELS 19 ANTWERP FLEMISH BRABANT WEST FLANDERS EAST FLANDERS HAINAUT LIEGE LIMBURG LUXEMBURG NAMUR
EDUCATIONAL LEVEL INTERVIEWED PERSON	NONE PRIMARY SECONDARY LOW SECONDARY HIGH GEN., TECHN. ART. SECONDARY HIGH PROF BACHELOR MASTER
PROFESSIONAL ACTIVITY INTERVIEWED PERSON	SELF EMPLOYED SALARY PUBLIC SECTOR SALARY PRIVATE SECTOR WITHOUT PROFESSIONAL ACTIVITY

PROFESSION INTERVIEWED PERSON	SMALL COMMERCE, ARTISAN, INDUSTRIAL AND FREELANCE 5-, FARMER UPPER MANAGEMENT, LIBERAL PROFESSIONS, BIG COMMERCE, INDUSTRIAL AND FREELANCE 6+ MIDDLE MANAGEMENT EMPLOYEE SKILLED WORKER UNSKILLED WORKER HOUSEWIFE RETIRED UNEMPLOYED STUDENT OTHER
MAIN RESPONSABLE FOR INCOME	MRI YES MRI NO
SOCIAL GROUPS	1   2   3   4   5   6   7   8   Don't know
NUMBER OF PEOPLE IN THE HOUSEHOLD	1   2   3   4   5+
GRANDPARENT	YES   NO
GRANDCHILDREN 0-2 YRS	YES   NO
GRANDCHILDREN 3-5 YRS	YES   NO
GRANDCHILDREN 6-11 YRS	YES   NO
GRANDCHILDREN 12-14 YRS	YES   NO
GRANDCHILDREN 15-17 YRS	YES   NO
GRANDCHILDREN 18-24	YES   NO
GRANDCHILDREN 25+	YES   NO
PERSON RESPONSABLE FOR PURCHASES	YES   NO
LANGUAGE MOST SPOKEN AT HOME	FRENCH DUTCH GERMAN ENGLISH OTHER LANGUAGE
INTERNET USAGE DEVICES	DESKTOP COMPUTER PORTABLE COMPUTER TABLET SMARTPHONE PORTABLE MULTIMEDIA CONSOLE TV PLAYER ANOTHER DEVICE
INTERVIEWED PERSON'S AGE	12-14   15-17   18-20   21-24   25-29   30-34 35-39   40-44   40-44   45-49   50-54   55-59 60-64   65-69   70-74   75+
INTERVIEWED PERSON AGE 18-	12-17
INTERVIEWED PERSON AGE 18+	18 +
INTERVIEWED PERSON'S AGE (7 cat.)	12-14   15-24   25-34   35-44   45-54   55-64   65+



PRESENCE OF CHILDREN	1 CHILD 2 CHILDREN 3 CHILDREN AND MORE NO CHILDREN -55 NO CHILDREN +55
PRESENCE CHILDREN -1 YEAR	YES   NO
PRESENCE CHILDREN 1 YEAR	YES   NO
PRESENCE CHILDREN 2 YEARS	YES   NO
PRESENCE CHILDREN 3-4 YEARS	YES   NO
PRESENCE CHILDREN 5-6 YEARS	YES   NO
PRESENCE CHILDREN 7-12 YEARS	YES   NO
PRESENCE CHILDREN 13-14 YEARS	YES   NO
5 Centers	YES   NO

## Annex 3: Calculation of social groups

Social groups are based on the educational level and occupation of the PRI (Person mainly Responsible for the Income).

For the CIM Internet study, we use a simplified version of the calculation of social groups in the CIM Golden Standard. Some versions of the popup questionnaire leave out the last occupation for people who are retired or jobless, and details on the PRI's occupation if the respondent is not the PRI.

Each educational level and occupation of the PRI receives a score. Both scores are then multiplied. The result is split into 8 groups, similar to the distribution of social groups in the internet population in the Virtual Population (see [www.cim.be/nl/golden-standard](http://www.cim.be/nl/golden-standard) for an explanation on the Golden Standard and the Virtual Population). Since the internet population has a higher social level than the general population, the highest social groups represent more than 1/8 of the panel.

The tables below show the scores used for education and occupation.

The values actually used as boundaries of the social groups may vary slightly from year to year.

Education	Score
primary school or no degree	10
lower general secondary education (first three years achieved)	35
lower secondary technical, artistic or professional education (first three years achieved)	25
higher general secondary education (last three years achieved)	50
higher secondary technical, artistic education (last three years achieved)	45
higher secondary professional education (last three years achieved)	40
candidate, bachelor (academic or professional), graduate	75
university license, master, post graduate, extra-university higher education (long type)	85
university license with additional degree, master after master	90
doctorate with thesis	100

Profession MRI	Score	Comment
Pupil / Student / in formation	10	
Suspension of professional occupation / time credit	50	
Housewife	10	
Incapacitated	10	
Jobless	29	60% of last profession (based on weighted average in VP2020)
Early retirement	42	75% of last profession (based on weighted average in VP2020)
Pensioned	34	60% of last profession (based on weighted average in VP2020)
Other	50	
Farmer	45	
Artisan	70	



Profession MRI	Score	Comment
Merchant, Industrialist 5-	70	
Merchant, Industrialist 6+	90	
Merchant, Industrialist	71	weighted average for merchant, industrialist (VP2020)
Skilled worker	50	
Unskilled worker	25	
Worker	46	weighted average for skilled and unskilled (VP2020)
Clerk with primarily office work	65	
Clerk with little or no office work	60	
Clerk	62	weighted average for office worker or not (VP2020)
Middle management 5-	70	
Middle management 6+	75	
Middle management	72	weighted average for middle management (VP2020)
Member of the general management, senior executive 5-	80	
Member of the general management, senior executive 6-10	90	
Member of the general management, senior executive 11+	100	
Member of the general management, senior executive	93	weighted average for general management (VP2020)
Free profession	100	
Freelance, self-employed 5-	70	
Freelance, self-employed 6+	90	
Freelance, self-employed	71	weighted average for freelance, self-employed (VP2020)

Age MRI
If Age MRI $\geq 45$ --> Score Education MRI * Score Profession MRI
If Age MRI $< 45$ --> Score Education MRI * Score Profession MRI - 1