# WHY 'SINGLE SOURCE' PANELS ARE NOT GOING TO PROVIDE GOOD ESTIMATES OF CROSS PLATFORM DUPLICATION PATTERNS AND WHAT CAN BE THE ALTERNATIVES?

## Peter Masson, Bucknull & Masson

*This paper is about producing reliable estimates of duplication between different types of content distribution channels. If these estimates are not reliable then everything that follows in terms of cross media 'package' or schedule evaluation is meaningless.*

*The paper discusses the difficulties with single source panels and with data fusion of independent surveys and suggests a different approach. This uses a large scale ad hoc 'Base Survey' to collect a limited amount of 'single source' cross media data. Each respondent record is then expanded and calibrated to simulate the results of an average week of data from each independent 'currency' survey in a process that leaves intact the initial 'single source' duplications however many integrations are involved.*

Finally cross channel schedule evaluation has some traction from media owners as they are being forced out of their silos and to deliver their content and 'package' advertising across platforms. They now have a 'need to know' the size and structure of their audiences, across time, which are common or different between distribution platforms for both ad.sales and content performance evaluation reasons.

This is not just an issue for print with digital platforms but also for video content providers. The TV screen will no longer be the preserve of 'broadcast' TV as new technology enables consumers to access competing Web Video on Demand content directly on their 'TV' screens. Not just for those with Smart TVs, but through inexpensive devices like the $35 Google Chromecast plug in. This will increasingly force broadcasters to distribute their program material via the Web in real time as VOD and not just as 'catch up', in an effort to expand/capture mobile audiences.

From a research perspective the 'crossover' between broadcast and broadband/mobile Web will place a much larger part of Web video and audio content streams onto the family 'TV' screen as a *communal viewing event,* rather than individual. This works in the TV homes PeopleMeter panel operator's favor where voice or tag recognition systems can potentially track the VOD content source and report the total reception and audience composition of these video and audio web streams.

But the *individual* video streaming of family members onto their other 'screens' will also need measurement and it is thus a natural extension to consider the development of an integrated single source 'homes' panel where video (and audio) content streaming is monitored on all devices used by family members to provide total audience and overlap data.

However recent presentations of such cross platform homes panels conducted in a number of countries have indicated the difficulties. Response rates were reported to be as low as 2% indicating that random probability sampling is not feasible and in most cases the control data they had available to restore the sample structure by weighting or by the use of quotas was very limited. In order to maintain a representative inter-device usage structure one would need a large scale and high response rate sample of homes (not individuals) against which to weight/control the sample/quotas. This would be using not only the normal TV panel homes selection/control variables (household composition by age within gender, social economic class, region, number of TV sets, type of signal reception, weight of TV viewing) but also to interlace these within (weight of) Web usage separately for PC/laptop, tablet and mobiles for each household member – and to maintain this sample structure in the light of panel drop out and continuing device growth. This is a near on impossible task.

We can observe the complexity of the inter-relationships of duplication that have to be maintained. The data in the following tables are taken from Orvesto Konsument 2012 which provides an unclustered 43,000 probability sample of the Swedish population with a response rate a little under 50%.

In the following two tables we consider the interactions of just 3 distribution channels: TV screen, PC/Laptop screen and radio which give rise initially to 7 interlaced duplication cells *(Radio/not TV/not I'Net;  Radio/I'Net;  Radio/TV/I'Net; I'Net/not TV/not Radio; TV/I'Net; TV/not I'Net/not radio; Radio/ TV)* but this is not sufficient for sample control. These 7 overlap areas need to be controlled further by the weight of usage of each medium as this correlates very highly with demographics and product sector interests/market behavior as the following two charts indicate. We have examined only three of the possible 50+ interlaced weight cells to demonstrate the differences.

| Table 1 Orvesto Consumer 2012:2 | | Heavy Radio /Light TV /Light I'Net | Heavy TV /Light Radio /Light I'Net | Heavy I'net /Light Radio /Light TV |
|---|---|---|---|---|
| Sample/% | 43035 | 678/1.6% | 4096/9.5% | 2864/6.7% |
| Population/% | 7354 | 122/1.7% | 556/7.6% | 605/8.2% |
| *Gender* | *C%^* | *Idx* | *Idx* | *Idx* |
| Men | 50.3 | 90 | 82 | 116 |
| Women | 49.7 | 110 | 118 | 84 |
| *Age Years* | | | | |
| 15 - 24 | 17 | 62 | 29 | 228 |
| 25 - 44 | 33.3 | 134 | 34 | 115 |
| 45 - 64 | 32.8 | 113 | 118 | 57 |
| 65 - 79 | 16.9 | 45 | 266 | 27 |

| Table 2 Orvesto Consumer 2012:2 | | Heavy Radio /Light TV /Light I'Net | Heavy TV /Light Radio /Light I'Net | Heavy I'net /Light Radio /Light TV |
|---|---|---|---|---|
| Sample/% | 43035 | 678/1.6% | 4096/9.5% | 2864/6.7% |
| Population/% | 7354 | 122/1.7% | 556/7.6% | 605/8.2% |
| *Rather-Very Interested in* | *C%^* | *Idx* | *Idx* | *Idx* |
| Motor Sport | 15 | 124 | 100 | 85 |
| Computers | 40.3 | 71 | 56 | 148 |
| New Tech In Everyday Life | 37.3 | 80 | 63 | 134 |
| Hair Care | 25.5 | 95 | 134 | 69 |
| Illness/Medical Treatment | 39.3 | 102 | 134 | 77 |
| Arts | 25.2 | 87 | 97 | 124 |
| Further Education/Courses | 35.9 | 98 | 61 | 141 |

The profiles of the 3 cells can be seen to be very different and it is essential for the correct evaluation of cross platform packages/schedules in a single source sample or in a sample where surveys have been integrated that these patterns are correctly represented.

And while sample structure is the primary issue for 'single source' panels the overall sample sizes achieved/afforded are also a problem. First, the duplications obtained from the panel minute-by-minute are wildly variable. Two adjacent minutes can easily give one duplication three or four times another. An average over all minutes can be dominated by one or two minutes. We are not at all confident that such duplications can be used successfully.

Second, the sample size of the panel survey cells is usually much too small to rely on the duplication accuracy, even if we were convinced that any meaningful average duplication could be obtained.

In essence this means that (Text) browsing, (Video) streaming behavior has to be measured from  (continuously reporting) samples of individuals representing each distribution channel separately (TV, PC/laptop, tablet and mobile) where it will be much easier to control and maintain the structure. A further independent ad hoc survey for (off line) print will also need to be present.

The TV PeopleMeter Panel can potentially track VOD content appearing on the TV screen. How current data owners react to allowing the Web Based newcomers access/participation in TV viewing 'currencies' will be very interesting and we may well see new panel data appearing outside the auspices of the current JICS and MOCS.
 But now of course we are left with the problem of *integrating* the different independent samples in a manner that can provide reliable estimates of the level and structure of duplication between individual vehicles delivered via each of the devices.

**Data Fusion**

Data fusion has been increasingly used to integrate these independent samples but difficulties are apparent in the control of the level and structure of duplication at the individual vehicle level and this of course is fundamental to any cross platform schedule or package evaluation. In fusion members of the two surveys are matched according to their closeness on the matching variables. But there is simply never enough sample to attempt to match on specific duplication patterns (the combination of which are huge) and of course the close matches would soon run out so that as the fusion progresses the closeness of the match can become very poor. At best therefore only the general weight of usage of a media type or group can be used in the matching process. The problems become greater as more media distribution types are to be included. In the end there is a danger that duplication estimates can end up little better than random.

There is also a further inherent problem for fusion in the media research environment. This is where it is required to match respondents from a clustered 'homes' sample, whose viewing and buying behavior is often dependent (on the behavior of other home members) with respondents from individual random samples where there is no such dependency. To match correctly (crucial for both viewing and target group behavior) would involve reconstructing the individual sample to an homes sample on the basis of the household composition data provided and then matching first within the constraints of the household structures.

We are therefore going to describe another form of data integration which we believe is a better alternative to Fusion for media research where duplication is such a crucial issue and where the 'homes' to individual matching issue is circumnavigated.

**The Calibrated VDiary**

The process is called Virtual Diary creation (VDiary[1]). In most countries we find a regular ad hoc print readership survey with extensive target group data based on a large scale quality sample of individuals. Generally it will also classify respondents 'single source' by weight and frequency of viewing TV, listening to radio (broadcast and web) and web browsing/streaming fixed and mobile - but not in sufficient detail for the reliable calculation of schedule/package evaluation statistics. We would normally use this as our '***Base Survey'.*** It is unnecessary to conduct a specific 'hub' survey (except for political reasons) as it simply increases the number of integrations necessary.

What the Virtual Diary process does is to take this limited 'single source' media information and 'expand' each informant's record as if he or she had ***completed a 7 day ¼ hour by ¼ hour diary*** such that the resulting reach and rating levels from summing all respondents ***provides the same reach and ratings levels found in the 'currency' panel or survey data***. Depending on the sample size of the panel from which the control data is derived the 'currency' reach and rating levels targets are set for a wide range of population subgroups designed to discriminate between different levels of viewing, browsing/streaming and listening. Generally we will make 15-30 separate population subgroups Vdiaries which are then combined to the total population.

This same expansion process can be applied to the data from each of the media distribution channels (Internet fixed and mobile, TV and Radio) available in the 'Base Survey'. ***This is a 'calibrated' (to currency) VDiary.***

We take the view that it is not necessary for the media sales 'packager' or agency planner to have the detailed week on week data from the panel in the cross-platform evaluation base. Rather they gain a better forecasting/planning perspective from working with an average week of data. It is not therefore our approach to attempt to integrate detailed long term (TV/radio/Internet) panel data, rather to provide an average week. Since the VDiary can be readily updated this average week can be re-created (calibrated) month to month or even weekly. Within the *Sesame* analysis this average week is modeled to 'n' weeks so that long term campaigns can be evaluated in a longitudinal manner.

**The Non Calibrated VDiary**

In some cases it is not required (or the data is not available as is the case in many emerging markets) to match the VDiary to the 'currency' survey data. In these cases we would produce a ***Non Calibrated VDiary***. The processes are very similar, although in a non-calibration situation we need more detailed data to be available on the 'Base Survey'. Further the diary that is created will represent an average week during the fieldwork period of the 'Base Survey'.

**The VDiary creation process**
The process has been described in some detail and evaluated in an earlier paper[2] but we will re-iterate some of the benefits and results from the way that duplication is managed in the diary creation process.

---

[1] © Sesame Systems Ltd.
[2] The better alternative to Fusion: A modeling procedure that simulates independent media 'currencies'.By Peter Masson and Dr. Paul Sumner, Sesame Systems Ltd., London, **ESOMAR Worldwide Multi Media Measurement Conference Shanghai, China, 4-7th June 2006**

For duplication we rely on guidance from the 'single source' TV/Radio/Internet fixed and mobile/Print questions which *are* asked on the 'Base Survey'. These usually include 'last week' and 'yesterday' questions and often 'yesterday day part' or 'yesterday quarter-hour'. Examples of the duplication control questions used are given below. In Table 3 at the week and day level for Orvesto Internet web and mobile sites (16 at present) and Table 4 at the day-part level but at the generic level.

**Table 3**



**Table 4**



With the aid of these questions we can represent very well the average duplication that exists. We accept that this is recall data and not a technical measurement but believe this to be a much more secure and stable estimate than any that can be derived from the panel survey.

In the process of calibration to the 'currency' control levels these claimed duplications are hardly disturbed. If the week reach level for a particular channel from the panel is lower than the 'Base Survey' week 'recency' claim we will randomly reject (within viewing frequency group) the required number of informants to meet the required panel reach level. If the panel reach is higher, then we will randomly select additional viewers for the diary week from those that viewed/visited in the last month and not the last week. The calibration required is generally limited. In the Swedish case (Table 5 below) the difference between the recall levels and the panel levels for Internet sites are not that great although there are notable differences between sites.

| Table 5 **Weekly reach '000** | Panel report | Orvesto recall | Index |
|---|---|---|---|
| Visit Aftonbladet.Se | 4611 | 4737 | 103 |
| Visit (Borås) Bt.Se | 119 | 118 | 100 |
| Visit Corren.Se | 154 | 166 | 108 |
| Visit Eniro.Se | 2629 | 2762 | 105 |
| Visit Expressen.Se | 2418 | 2385 | 99 |
| Visit Gp.Se | 601 | 513 | 85 |
| Visit Hitta.Se | 3830 | 3006 | 78 |
| Visit Sf.Se | 432 | 862 | 200 |

This duplication decision to use the 'Base Survey' recall data has consequences when we come to compare a *Sesame* schedule evaluation using the calibrated survey against an average of a number of panel runs. We expect and find Gross statistics to be very similar between the two methods but Net results may vary as the results from the panel will be based on actual minute by minute spots transmissions while those from the VDiary are based on an average minute in the segment in an average week.  However in situations where we are able to test the results are always very consistent with the panel schedule evaluation results even at the 5+ reach levels as the following Swedish example indicates*:

| Soya: 550 spot campaign over 2 months with 6 channels | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Table 6** | | Gross Reach% MMS Panel | Sesame VDiary | 1+ Reach% MMS panel | Sesame VDiary | 5+ Reach% MMS panel | Sesame D |
| Adults | 15-79 | 440 | 452 | 71.6 | 70.1 | 32.6 | 32 |
| Men | 15+ | 437 | 461 | 71.2 | 69.5 | 32.6 | 32.4 |
| Women | 15+ | 443 | 442 | 72 | 70.6 | 32.7 | 31.6 |
| Adults | 15-29 | 446 | 447 | 58.4 | 63.7 | 31.4 | 30.8 |
| Adults | 30-49 | 481 | 465 | 76.5 | 70.3 | 35.4 | 32.2 |
| Adults | 50+ | 398 | 445 | 74.2 | 73.6 | 30.1 | 32.4 |

Sources: Orvesto Kosument 2011.3 and MMS June July 2011

* VDiaries – calibrated or not – are created originally for TV/Radio/Internet for a one week period. Our planning software *Sesame* provides a model to extend the VDiary week to 'n' weeks so that the planner can evaluate a schedule over any time period. This very complex model maintains the time dynamics of each of the different media.  100 TV rating points or 1 million Page Views placed in 3 days will exhibit a different net reach and OTS distribution than if placed over 7 or 14 days, just as would be found by counting the results on a continuous (longitudinal) panel. This indeed is one of the most important aspects of the VDiary and its application in the Sesame software is that it simulates longitudinal data.

The fact that the VDiary process can simulate very well the 'currency' survey in terms of schedule Reach and Frequency results within a media type is a strong indicator it will do so across media types since the 'single source' duplication control process is identical within and between media groups.

**Cross media analyses**

Below we present some results to indicate the metrics that are available from integrated VDiary analysis (Survey Base Orvesto Consumer 2012.3 QRS):
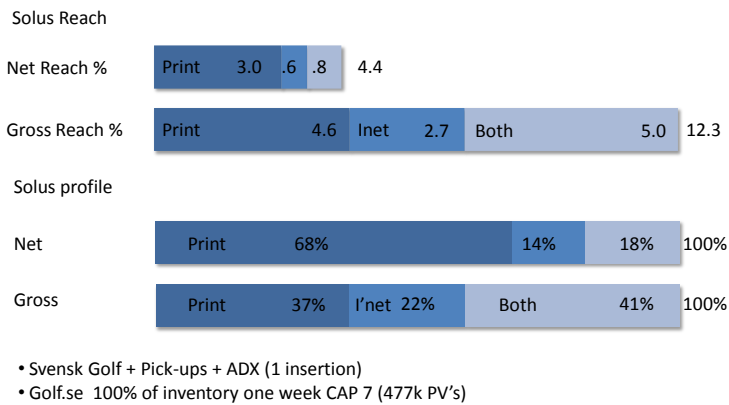
**Table 7**
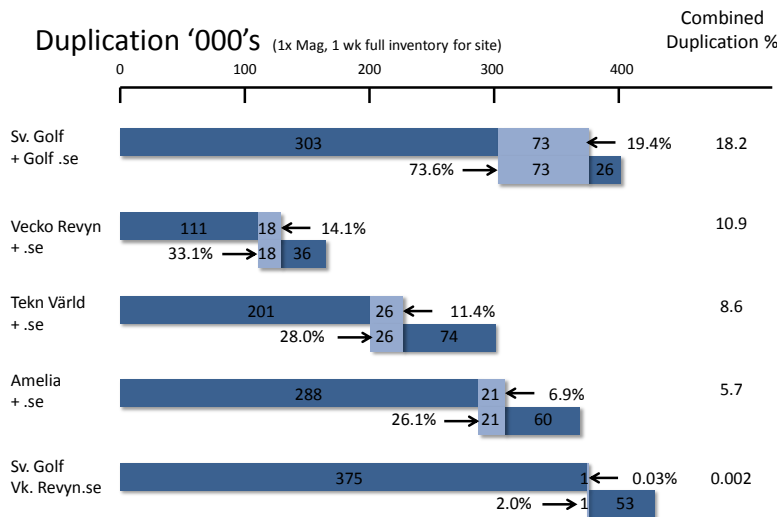


This is a straightforward duplication analysis using 1 insertion in the magazine and the one week full inventory 'buy' in the site of the magazine.  The analysis can also be made using proportions of the full inventory and where the ad. servings are capped to given OTS levels.  The light blue areas represent the number exposed to both the site and the paper publication.  In the case of Svensk Golf 19.4% of the readers also had an on-line exposure to Golf.se.  Of those that were exposed to Golf.se online 73.6% also read Svensk Golf. This is a combined duplication of 18.2%.  The different and declining relationship between site and title can be seen with the extreme situation between the disparate golf title (Svensk Golf) and the Vecko Revyn.se site (old established women's magazine) with virtually no duplication.  The VDiary will be reflecting the actual observed duplication in the single source 'Base Survey'.

**Table 8**

### 'Reach Contribution*' to 'Ad. Package'

Solus Reach

| | | | |
|---|---|---|---|
| Net Reach % | Print 3.0 | .6 .8 | 4.4 |
| Gross Reach % | Print 4.6 | Inet 2.7 | Both 5.0 | 12.3 |

Solus profile

| | | | |
|---|---|---|---|
| Net | Print 68% | 14% | 18% | 100% |
| Gross | Print 37% | I'net 22% | Both 41% | 100% |

• Svensk Golf + Pick-ups + ADX (1 insertion)
• Golf.se 100% of inventory one week CAP 7 (477k PV's)

In the table above we can see the reach contribution from each of the media components in a schedule or package. In this case it was a simple package/schedule consisting of 1 insertion in Svensk Golf and the full inventory for one week in the Golf.se site.

However even in such a simple evaluation we run into areas of media exposure 'equivalence' and this is particularly the case in a print/Internet evaluation. In print the Recent Reading model does not report multiple reading events or ad page exposure. In the case of Internet each Page View 'clocks up' up a gross contact to the page on which the advertisement is served (even if below the scroll).
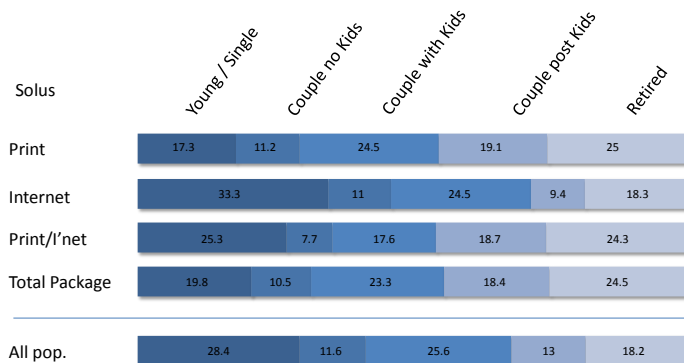
Within the Orvesto Survey we have integrated additional print reading information allowing multiple pick ups to be included in the gross count for magazines but at the same time discounting exposure on the basis of the proportion of issue read. These measures are included for Svensk Golf in the above analysis. At the same time while we have 'bought' the total inventory of Golf.se we have restricted (capped) ad servings to a browser to 7 during the week.

The addition of Golf.se (under the above conditions) increases Net reach from 3.0% to 4.4% (+47%). Those who only read the print edition account for 3% points of this, those who only visit the site account for 0.6% and for those that both read and browse contribute 0.8% to reach. Looked at in profile terms 68% of the reach is provided by those only reading the print edition, 14% by those only visiting the site and 18% who do both.

The Gross Reach situation is very different, it increases from 4.6% to 12.3% (+267%). Here solus print readers account for only 37%, solus Internet site visitors 22% and those who do both 41%. It becomes quite clear that the composition of the Net audience will be very different from the composition of the Gross audience and we go further to examine these for the same schedule/package conditions below:
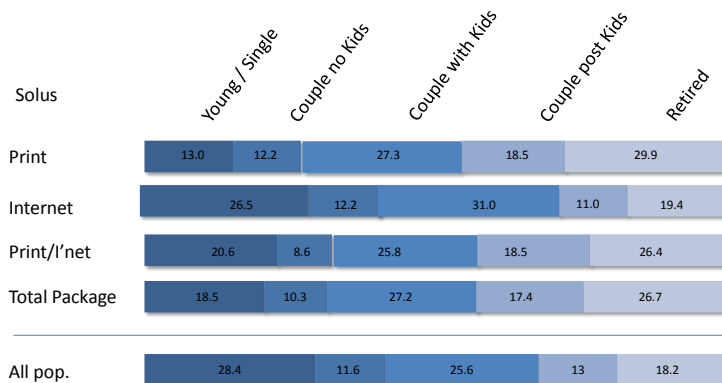
**Table 9**

### 'Ad. Package' Net Reach profiles

| Solus | Young / Single | Couple no Kids | Couple with Kids | Couple post Kids | Retired |
|---|---|---|---|---|---|
| Print | 17.3 | 11.2 | 24.5 | 19.1 | 25 |
| Internet | 33.3 | 11 | 24.5 | 9.4 | 18.3 |
| Print/I'net | 25.3 | 7.7 | 17.6 | 18.7 | 24.3 |
| Total Package | 19.8 | 10.5 | 23.3 | 18.4 | 24.5 |
| All pop. | 28.4 | 11.6 | 25.6 | 13 | 18.2 |

**Table 10**

## 'Ad.Package' Gross Reach profiles

| Solus | Young / Single | Couple no Kids | Couple with Kids | Couple post Kids | Retired |
|---|---|---|---|---|---|
| Print | 13.0 | 12.2 | 27.3 | 18.5 | 29.9 |
| Internet | 26.5 | 12.2 | 31.0 | 11.0 | 19.4 |
| Print/I'net | 20.6 | 8.6 | 25.8 | 18.5 | 26.4 |
| Total Package | 18.5 | 10.3 | 27.2 | 17.4 | 26.7 |
| | | | | | |
| All pop. | 28.4 | 11.6 | 25.6 | 13 | 18.2 |

Once one introduces multiple 'pick ups' for print and Page View for Internet the frequency of visit aspects come into play in the Gross profile which may then be quite different from the Net profile and this is demonstrated in the examples above. The solus Gross Print profile shows a notable increase in the proportion of 'couples with kids' and 'retired's', while the solus Internet and combined Print/Internet profiles also age more and very largely to the 'couple with kids' sector. What is clear however, is that the VDiary construction retains very well the subtle differences in profiles within duplicated audiences.

## Concluding remarks

We use of a large scale ad hoc 'Base Survey' to provide 'single source' duplication patterns vehicle by vehicle within multiple media types. The data record for the reading/viewing/browsing for each respondent is 'expanded' to simulate their responses to a full 7 day diary week (1/4 hour by 1/4 hour) such that the totality of their responses reproduces well the reach, rating and duplication patterns of independent 'currency' surveys (be they ad hoc or panel). This modeling and calibration process is the Virtual Diary creation process.

In data integration processes there is always some element of regression to the mean. However the evidence, by now quite extensive, that we have is that our calibration methods are much more stable in this respect than any comparable fusion method. This is because, towards the latter part of any fusion technique, the matching of respondents is often rather poor. The calibration matching procedure, dependent as it is on an unambiguous cell matching, is strong and consistent from beginning to end of the matching. Also in calibration/VDiary creation we proceed title by title, and slot by slot, and so can ensure that duplication is very well controlled at the weekly and daily reach level. This then carries through rather well to individual slots, and the duplication between slots. Fusion methods cannot exert this control over individual titles/slots.

It is also unwise to compare duplications between say two panel weeks and the average produced by calibration. The difference between two or several panel weeks are always much larger than the difference between any one panel week and the calibrated average. The calibrated result is a much more reliable predictor of performance than any single panel week.
These issues become particularly important when we are attempting to unite three or more surveys, say Press, TV, Internet and Radio. The regression to the mean and impossibility of duplication control of fused surveys is then so extreme as to render the fusions very unreliable indeed. Calibration/VDiary creation becomes by far the more reliable method.