



---

## **Deliverable 1.3: Updated data, quality and knowledge management plan**

---

E. Apostolidis, S. Papadopoulos, M. Zampoglou, V. Mezaris (CERTH), D. Teyssou (AFP), R. Garcia (UdL), L. Nixon (MODUL), A. Scharl, M. Göbel (WLT), G. Rudinger (APA-IT), T. Koch (DW), R. Fricke (Condat), G. Mercier (EXO)

30/09/2017

Work Package 1: Project and Innovation Management

## **InVID - In Video Veritas: Verification of Social Media Video Content for the News Industry**

Innovation Action

Horizon 2020, Research and Innovation Programme

Grant Agreement Number 687786

Dissemination level	<i>Public</i>
Contractual date of delivery	<i>30/09/2017</i>
Actual date of delivery	<i>30/09/2017</i>
Deliverable number	<i>D1.3</i>
Deliverable name	<i>Updated data, quality and knowledge management plan</i>
File	<i>InVID_D1.3_v1.0.pdf</i>
Nature	<i>Report</i>
Status & version	<i>Final, v1.0</i>
Number of pages	<i>43</i>
WP contributing to the deliverable	<i>WP1</i>
Task responsible	<i>CERTH</i>
Other contributors	<i>AFP, UdL, MODUL, WLT, APA-IT, DW, Condat, EXO</i>
Author(s)	<i>Evlampios Apostolidis, Symeon Papadopoulos, Markos Zampoglou, Vasileios Mezaris (CERTH), Denis Teyssou (AFP), Roberto Garcia (UdL), Lyndon Nixon (MODUL), Arno Scharl, Max Göbel (WLT), Gerhard Rudinger (APA-IT), Tim Koch (DW), Rolf Fricke (Condat), Gregoire Mercier (EXO)</i>
Quality Assessor	<i>Rolf Fricke (Condat), Jochen Spangenberg (DW)</i>
EC Project Officer	<i>Miguel Montarelo-Navajo</i>
Keywords	<i>InVID Data Management Plan, InVID Quality Management Plan, InVID Knowledge Management Plan</i>

**Abstract:**

This deliverable documents the updated data, quality and knowledge management plans of the InVID project. It updates the Data Management Plan presented in D1.1, by introducing new datasets and reporting on datasets that are no longer relevant to the needs of the project. Moreover, it discusses the outcomes of the Quality Management Plan described in D1.2 after being applied over the first 21 months of the project and provides updates on aspects of this plan. Finally, it extends the list of Intellectual Property assets of the InVID project that are part of the applied Knowledge Management Plan described in D1.2. Through this reporting, the deliverable outlines the experience gained so far regarding data, quality and knowledge management.

## Table of contents

<b>1</b>	<b>Introduction.....</b>	<b>5</b>
1.1	History of the document .....	5
<b>2</b>	<b>Updated Data Management Plan .....</b>	<b>6</b>
2.1	Datasets no longer relevant .....	6
2.2	Datasets newly created or updated .....	6
<b>3</b>	<b>Updated Quality Management Plan .....</b>	<b>18</b>
3.1	Innovation management.....	18
3.2	Research activities .....	20
3.3	Software development .....	20
3.4	Project deliverables.....	22
3.5	Overall project progress and compliance with the timeplan .....	23
<b>4</b>	<b>Updated Knowledge Management Plan .....</b>	<b>27</b>
4.1	WP2 IP assets .....	27
4.2	WP3 IP assets .....	30
4.3	WP4 IP assets .....	34
4.4	WP5 IP assets .....	36
4.5	WP6 IP assets .....	38
4.6	WP7 IP assets .....	40
<b>5</b>	<b>Summary .....</b>	<b>43</b>

# 1 Introduction

This deliverable documents the updated data, quality and knowledge management plans, based on the experience gained during the first 21 months of the project. The first part of the document (Section 2) provides an update of the first release of the Data Management Plan presented in D1.1. Following the same methodology and template, it reports on datasets that are no longer relevant to the software development activities of the project (also explaining why) and presents newly-introduced datasets after the first release of the Data Management Plan. The second part of the document (Section 3), discusses the outcomes of the applied Quality Management Plan reported in D1.2, and provides updates on this plan, regarding the innovation management, the research activities, the software development, the project deliverables, the overall progress of the project and its compliance with the predefined timeplan. The third part of the document (Section 4) updates and enriches the list of Intellectual Property assets of the InVID project, given the fact that the Knowledge Management Plan presented in D1.2 is still fully applicable. The final Section 5 provides a summary of the document.

## 1.1 History of the document

**Table 1: History of the document**

Date	Ver.	Name	Comment
02/08/2017	V0.1	E. Apostolidis, V. Mezaris (CERTH)	Structure, writing guidelines and writing assignments
07/09/2017	V0.2	M. Zampoglou, S. Papadopoulos (CERTH), R. Garcia (UdL), R. Fricke (Condat), T. Koch (DW), G. Rudinger (APA-IT), D. Teyssou (AFP)	Contributions on WP3, WP4 datasets and WP3, WP4, WP6, WP7, IP assets. New innovation management part
13/09/2017	V0.3	L. Nixon (MODUL)	Contributions on WP2, WP3 datasets and IP assets.
19/09/2017	V0.4	E. Apostolidis, V. Mezaris (CERTH)	Prepared other parts of the document
20/09/2017	V0.5 V0.6	E. Apostolidis (CERTH), G. Mercier (EXO), A. Scharl, M. Göbel (WLT)	Integrated WLT and EXO IP assets and submitted to QA
27/09/2017	V0.7	E. Apostolidis (CERTH)	Newer version based on reviewers' feedback
30/09/2017	V1.0	E. Apostolidis, V. Mezaris (CERTH)	Final version after a last check submitted to the EC.

## 2 Updated Data Management Plan

This section updates the Data Management Plan of the InVID project, which was originally presented in D1.1. Following the same template and methodology, it focuses on datasets that are no longer relevant (also justifying why) and newly-introduced or updated datasets during the last 17 months of the project's life.

### 2.1 Datasets no longer relevant

Datasets described in D1.1 that are no longer relevant (have not been used and will not be used in the project) include the following:

**InVID\_Data\_WP3\_1\_WildWebTamperedImages:** While the initially proposed approach entailed an evaluation of the TUNGSTEN forensics algorithms on the Wild Web Tampered Images dataset, the forensics investigations of WP3 revealed that video forensics are radically different from image forensics, and that the problem cannot be reduced to applying existing image forensics algorithms on extracted keyframes. Therefore, image forensics algorithms are not particularly relevant to the task, and instead of evaluating image forensics algorithms on the Wild Web dataset, it was decided to divert effort towards collecting video content for evaluation.

**InVID\_Data\_WP3\_3\_VisualGeometryGroupDatasets:** This dataset was intended to be used for evaluating near-duplicate detection (NDD) methods that were based on detecting near-duplicity relations among images/video keyframes. However, the main near-duplicate video detection method that was implemented in the first year was applicable on whole videos. As a result, this dataset is no longer relevant to our evaluations. Instead, we are developing our own video dataset (named "InVID\_Data\_WP3\_9\_NearDuplicateNewsVideos") for that purpose.

**InVID\_Data\_WP3\_4\_InriaDatasets:** For the same reasons as the ones that led to "InVID\_Data\_WP3\_VisualGeometryGroupDatasets" being removed from our list, this image dataset is no longer relevant for our evaluations.

### 2.2 Datasets newly created or updated

Datasets that have been created or updated after releasing the first version of the InVID Data Management Plan on April of 2016, are the following:

Dataset name	<b>InVID_Data_WP2_5_NUS_WIDE</b>
Dataset description	NUS-WIDE is a web image dataset created by the Lab for Media Search at the National University of Singapore. It includes: 269,648 images and the associated tags from Flickr, with a total number of 5,018 unique tags; 6 types of low-level features extracted from these images, including 64-D color histogram, 144-D color correlogram, 73-D edge direction histogram, 128-D wavelet texture, 225-D block-wise color moments and 500-D bag of words based on SIFT descriptions; and ground-truth for 81 concepts that can be used for evaluation. It will be used for fine-tuning ImageNet Deep Convolutional Neural Networks (DCNNs) for video concept detection.
Standards and metadata	This static dataset is composed by images that are mainly in high resolution JPEG format. The created metadata, after analyzing these images, can be: (a) local features extracted from these images that are stored in BIN or TXT files, and (b) the output of the trained DCNNs (i.e. the classification decision), which is stored in TXT files. These data will be accompanied by a document (a word file) containing metadata with sufficient information to: (a) link it to the research publications/outputs, (b) identify the funder and discipline of the research, and (c) help internal users to locate the data.
Data sharing	The NUS-WIDE dataset is freely available for non-commercial research and/or educational purposes, after following the suggested procedure and adopting the terms of use that are described in the NUS-WIDE website <sup>1</sup> .
Archiving and preservation	The original dataset and the results of processing it will be stored on the file servers of CERTH (protected by applying the commonly used security measures for preventing unauthorized access and ensuring that security software is up-to-date with the latest released security patches) and backup provisions will be made. The archiving and preservation of this dataset are performed by the National University of Singapore; InVID will have no involvement in this process.

<sup>1</sup> <http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm>

Dataset name	<b>InVID_Data_WP2_6_ConceptDetectionScoresAVS</b>
Dataset description	<p>This dataset contains the concept detection scores for the IACC.3 dataset (600 hrs internet archive videos), which is used in the TRECVID Ad-hoc Video Search (AVS) task<sup>2</sup>. Concept detection scores for 1,345 concepts (1,000 ImageNet concepts provided for the ILSVRC challenge<sup>3</sup> and 345 TRECVID SIN concepts<sup>4</sup>) have been generated as follows:</p> <ul style="list-style-type: none"> <li>• for the ImageNet concepts, 5 pre-trained ImageNet networks were applied on the IACC.3 dataset and their output was fused in terms of arithmetic mean;</li> <li>• for the TRECVID SIN concepts, two pre-trained ImageNet networks were fine-tuned on these concepts using a combination of our methods<sup>5,6</sup>.</li> </ul> <p>The dataset contains two different sets of concept scores for the TRECVID SIN concepts:</p> <ul style="list-style-type: none"> <li>• the output of the two fine-tuned networks was fused in terms of arithmetic means in order to return a single score for each concept;</li> <li>• the last fully-connected layer was used as a feature to train SVM classifiers separately for each fine-tuned network and each concept; then, the SVM classifiers were applied on the IACC.3 dataset and the prediction scores of the SVMs for the same concept were fused in terms of arithmetic mean in order to return a single score for each concept.</li> </ul> <p>This dataset can be used for reproducing and validating the results of the developed InVID algorithms for concept detection that were reported in the</p>

<sup>2</sup> G. Awad, J. Fiscus, M. Michel et al. 2016. TRECVID 2016: Evaluating Video Search, Video Event Detection, Localization, and Hyperlinking. In TRECVID 2016 Workshop. NIST, USA.

<sup>3</sup> O. Russakovsky, J. Deng, H. Su et al. 2015. ImageNet Large Scale Visual Recognition Challenge. Int. Journal of Computer Vision (IJCV) 115, 211–252.

<sup>4</sup> G. Awad, C. Snoek, A. Smeaton, and G. Quénot. 2016. TRECVID semantic indexing of video: a 6-year retrospective. ITE Transactions on Media Technology and Applications, 4 (3). pp. 187-208.

<sup>5</sup> N. Pittaras, F. Markatopoulou, V. Mezaris, I. Patras. 2017. Comparison of Fine-tuning and Extension Strategies for Deep Convolutional Neural Networks, Proc. 23rd Int. Conf. on MultiMedia Modeling, Reykjavik, Iceland, Springer LNCS vol. 10132, pp. 102-114, Jan. 2017.

<sup>6</sup> F. Markatopoulou, V. Mezaris, I. Patras. 2016. Deep Multi-task Learning with Label Correlation Constraint for Video Concept Detection, Proc. ACM Multimedia 2016, Amsterdam, Oct. 2016.

Dataset name	<b>InVID_Data_WP2_5_NUS_WIDE</b>
	aforementioned scientific publications.
Standards and metadata	<p>Three different files of concept detection scores can be downloaded (after unpacking the compressed file):</p> <p>1) scores_ImageNet.txt</p> <p>2a) scores_SIN_direct.txt</p> <p>2b) scores_SIN_svm.txt</p> <p>In total there are 335,944 rows in each file; 1,002 columns in the first file and 347 columns in each of the other two. Each row in any of these files corresponds to a different video shot; the video shot IDs appear in the first two columns. (Note: the shot IDs are the ones from the mp7 files in the TRECVID AVS master shot reference, with the format "shotFILENUMBER_SHOTNUMBER"). Then, each column (except for the first two) corresponds to a different concept, with all concept scores being in [0,1] range. The higher the score the more likely that the corresponding concept appears in the video shot. Files "concept_names_ImageNet.txt" and "concept_names_SIN.txt" indicate the order of the concepts that is used in the concept score files.</p>
Data sharing	The dataset and its documentation are publicly accessible through the InVID community on Zenodo <sup>7</sup> .
Archiving and preservation	The dataset is already archived on Zenodo, and a copy is maintained in CERTH and used for further evaluations.

<sup>7</sup> <https://zenodo.org/record/292994>

Dataset name	<b>InVID_Data_WP3_2_FakeVideoCorpus (UPDATED)</b>
Dataset description	This dataset was first described in D1.1 as a future plan, containing a list of known fakes to be used as a benchmark for verification. The dataset has now been collected, a first version containing 59 fakes has been presented in D3.1 and published on Zenodo, and further updates are underway.
Standards and metadata	For each video, information is provided describing the fake, its original source, and the evidence proving it is a fake. As we do not own the videos, the dataset only provides the video URLs and metadata, in the form of a tab-separated value (TSV) file.
Data sharing	The first version of the dataset and the corresponding documentation has already been published openly at the InVID Community on Zenodo <sup>8</sup> . The metadata are openly distributed, while the videos themselves are property of YouTube.
Archiving and preservation	Besides the version already published, a running version is maintained by CERTH and continuously enriched with new videos. In parallel, a second set of real user-generated videos has already been collected. The aim is to provide a benchmark for video verification that can evaluate the ability of algorithms to identify fakes while simultaneously avoiding false positives. An updated version of the dataset will thus be presented in D3.2, also to be archived in the InVID community on Zenodo.

Dataset name	<b>InVID_Data_WP3_8_NewsLogos (UPDATED)</b>
Dataset description	This dataset was first reported in D1.1 under the name "TVChannelsLogos", as a future plan aimed to provide an annotated collection of videos containing logos from TV channels and user-generated channels. This dataset was designed to provide an evaluation benchmark

---

<sup>8</sup> <https://zenodo.org/record/242481>

Dataset name	<b>InVID_Data_WP3_8_NewsLogos (UPDATED)</b>
	<p>for the logo detection module of T3.2.</p> <p>As presented in D3.1, this dataset was developed by CERTH and is now complete. It contains the results from the segmentation and annotation of 2,749 YouTube videos originating from a large number of TV news channels. The videos have been annotated with respect to the TV channel logos they contain - specifically, by the name of the organization to which the logo belongs - and with shot boundary information. Furthermore, a set of logo templates has been extracted from the videos and organized alongside the corresponding channel information. Currently, the dataset is being expanded with additional logos from various groups and factions from the Syrian Civil War, as well as other informal groups whose logos may appear in news-related user-generated videos. Thus, the name is changed to reflect this broadening in the dataset's scope.</p>
Standards and metadata	As we do not own the rights to the videos, the dataset only contains the YouTube video IDs alongside the corresponding annotations. It further contains 503 logo template files and the corresponding metadata information (channel name, Wikipedia link). The information is stored in tab-separated value (TSV) format.
Data sharing	The dataset and the corresponding documentation has already been published openly on Zenodo <sup>9</sup> . The metadata are openly distributed, while the videos themselves are property of YouTube.
Archiving and preservation	The dataset is already archived on Zenodo, and a copy is maintained in CERTH and used for further evaluations.

---

<sup>9</sup> <https://zenodo.org/record/240825>

Dataset name	<b>InVID_Data_WP3_9_NearDuplicateNewsVideos</b>
Dataset description	This dataset is under development as part of task T3.2, aiming at the detection of near-duplicate content among videos originating from mainstream video platforms. To create this corpus, we are collecting all events that took place from January 2013 to June 2017 by crawling Wikipedia's "Current Events" tab <sup>10</sup> , and we are filtering them based on their topic. The headlines of the retained events are provided as queries to the public APIs of social media video platforms (i.e. Youtube, Dailymotion) and news agencies (i.e. The Guardian, New York Times) to collect a set of video URLs which will form the final corpus. The dataset will be manually annotated based on the similarity of the content between videos.
Standards and metadata	The dataset is going to contain the video URLs, the video keyframes and visual descriptors of the collected videos. Also, an annotation scheme will be defined to accommodate the needs of this corpus. The serialization format of the metadata will most likely be JSON to enable easy parsing, extensibility and ease of storage and retrieval.
Data sharing	Since the corpus will be collected by the InVID consortium, we plan to make the dataset publicly available through the InVID community on Zenodo. However, since part of the data will come from third party platforms (e.g. YouTube, Dailymotion, etc.), we cannot openly distribute the original videos due to legal restrictions. Instead, we will provide video URLs, video keyframes and visual descriptors.
Archiving and preservation	We intend to archive the dataset on Zenodo. Backup provisions and the results of processing will be stored securely on the file servers of CERTH.

---

<sup>10</sup> [https://en.wikipedia.org/wiki/Portal:Current\\_events](https://en.wikipedia.org/wiki/Portal:Current_events)

Dataset name	<b>InVID_Data_WP3_10_StoryLens</b>
Dataset description	This a dataset focused on geolocation detection and events. It contains around 400 documents from the summer of 2017 split in 3 partitions from the following streams: news, YouTube and Twitter. The dataset was created specifically to highlight the difficulties of identifying correct locations in news and social media data streams, and includes multiple types of annotations: a) the classic Location, Facility, Geo Political Entity, Person, Organization; b) events and stories. DBpedia links are used for the entities. The annotation guideline is based on TAC-KBP and ACE guidelines. The dataset will be available in December 2017.
Standards and metadata	The data of this dataset are stored in Natural Language Processing Interchange Format (NIF) and Comma-Separated Value (CSV) format.
Data sharing	Licensed under a Creative Commons license (by-nc-sa 4.0).
Archiving and preservation	The dataset will be archived and preserved on the MODUL Technology Github page and the InVID community on Zenodo.

Dataset name	<b>InVID_Data_WP3_11_N3Collection</b>
Dataset description	This is a collection of 3 datasets (Reuters128, News100, RSS500) created specifically for NEL (Named Entity Linking) in 2014 <sup>11</sup> . The datasets were extracted from the classic Reuters corpus, news media and RSS feeds and were published by AKSW/DICE group. DBpedia links are used for the entities. These datasets are general purpose, but since they are news-

---

<sup>11</sup> Described in the paper: "Michael Röder, Ricardo Usbeck, Sebastian Hellmann, Daniel Gerber, Andreas Both: N<sup>3</sup> - A Collection of Datasets for Named Entity Recognition and Disambiguation in the NLP Interchange Format. LREC 2014: 3529-3533"

Dataset name	<b>InVID_Data_WP3_11_N3Collection</b>
	related and from different time periods (mid-80s to this decade) we decided to use them for geolocation and event detection.
Standards and metadata	The data of this dataset are stored in Natural Language Processing Interchange Format (NIF).
Data sharing	Licensed under a Creative Commons license (by-nc-sa 4.0).
Archiving and preservation	This dataset is archived on GitHub at: <a href="https://github.com/dice-group/n3-collection">https://github.com/dice-group/n3-collection</a> .

Dataset name	<b>InVID_Data_WP3_12_EyeOfTheHurricane</b>
Dataset description	This is a dataset that contains toponym annotations based on a set of tweets that happened during a Hawaii hurricane <sup>12</sup> . It is used for improving and evaluating the performance of the InVID method for location detection.
Standards and metadata	The data of this dataset are stored in General Architecture for Text Engineering XML (GATE XML) and Comma-Separated Value (CSV) format.
Data sharing	Unspecified license, but free for research purposes.

---

<sup>12</sup> Andrea Salfinger, Caroline Salfinger, Birgit Pröll, Werner Retschitzegger, Wieland Schwinger. Pinpointing the Eye of the Hurricane - Creating a Gold-Standard Corpus for Situative Geo-Coding of Crisis Tweets Based on Linked Open Data. Proceedings of the LREC 2016 Workshop "LDL 2016 - 5th Workshop on Linked Data in Linguistics: Managing, Building and Using Linked Language Resources". John P. McCrae, Christian Chiarcos et al., editors, pages: 27 - 35. May, 2016.

Dataset name	<b>InVID_Data_WP3_12_EyeOfTheHurricane</b>
Archiving and preservation	<p><a href="https://weizenbaum.tk.i.ku.at/owncloud/public.php?service=files&amp;t=6076c0c9b7f3e03fc6204b1607a8b0e1">https://weizenbaum.tk.i.ku.at/owncloud/public.php?service=files&amp;t=6076c0c9b7f3e03fc6204b1607a8b0e1</a></p> <p>We have translated links to DBpedia, but do not have any plans to republish this dataset. We might consider publishing the scripts that translated the links.</p>

Dataset name	<b>InVID_Data_WP4_2_UGCReuseAgreements</b>
Dataset description	<p>This dataset contains the agreements between user-generated content (UGC) reusers, usually journalists, and UGC owners, usually the social networks' users that uploaded them. These agreements are also stored in the dataset "InVID_Data_WP4_1_UGCRegisteredProviders", with additional information about content providers or reuse policies. Agreements are also stored in this dataset in order to timestamp them. Moreover, as this dataset will be made persistent using immutable storage technologies<sup>13</sup>, agreements will also be tamper-proof. We will use this for the agreements, which will be stored encrypted to preserve privacy. Consequently, the main objective of this dataset is to store the agreements between content owners and reusers in a way that they serve as proofs.</p>
Standards and metadata	<p>The dataset is based on blockchain technologies, namely the Ethereum blockchain. The agreements will be based on the Copyright Ontology, just like "InVID_Data_WP4_1_UGCRegisteredProviders", but serialized using JSON-LD. This information is then stored in the blockchain using a canonical binary form based on RFC 7049, the Concise Binary Object Representation.</p>
Data	<p>This dataset is generated as a result of the InVID platform operation when</p>

---

<sup>13</sup> When something is stored in this kind of distributed ledger it cannot be changed. Any attempt for modification would be detected thanks to the tamper-proof property of this technology.

Dataset name	<b>InVID_Data_WP4_2_UGCReuseAgreements</b>
sharing	the Rights Module is involved and is specific to its operation. This dataset basically contains bilateral agreements between content owners and the reusers, which we assume that they will prefer not to fully expose in public. However, data on a public blockchain, like Ethereum, is shared across all nodes participating in the blockchain and it is thus publicly available. Consequently, what is stored in the blockchain is just a hash of the agreement and the details are encrypted.
Archiving and preservation	This dataset will be preserved at the same location where the Rights Management module is deployed, i.e. a server hosted at the premises of Universitat de Lleida as detailed in the Privacy Policy included in D4.2. It will be protected by preventing unauthorized access to the server and ensuring that security software is up-to-date. Moreover, backup provisions will be made.

Dataset name	<b>InVID_Data_WP8_3_InVidPublications (UPDATED)</b>
Dataset description	This dataset contains manuscripts reporting the conducted scientific work in InVID that have been accepted for publication in peer-reviewed journals and conferences. All these publications include a statement with acknowledgement to the InVID project, while their content may vary from the description of specific analysis techniques, to established evaluation datasets and individual components or parts of the InVID platform.
Standards and metadata	Most commonly, these documents are stored in PDF format. Each document is accompanied by: (a) details about the venue (e.g. conference, workshop or benchmarking activity) or journal where it was published, (b) a short description with the abstract of the publication, and (c) the LaTeX-related BIB file with its citation. This dataset is extended whenever new submitted works are accepted for publication in conferences or journals. A simple log file of the performed updates of the dataset is maintained by CERTH in the project wiki (hosted by a CERTH server).

Dataset name	<b>InVID_Data_WP8_3_InVidPublications (UPDATED)</b>
Data sharing	This dataset is publicly available via the InVID Community on Zenodo ( <a href="https://zenodo.org/communities/invid-h2020">https://zenodo.org/communities/invid-h2020</a> ).
Archiving and preservation	Self-archiving (also known as "green" open access) is applied for ensuring open access to these publications. According to this archiving policy the author(s) of the publication archive (deposit) the published article or the final peer-reviewed manuscript in online repositories, such as personal webpage(s), the project website <sup>14</sup> and the free-of-charge Zenodo repository, after its publication. Nevertheless, the employed archiving policy is fully aligned with restrictions about embargo periods that may be defined by the publishers of these publications, making the latter publicly available in certain repositories only after an embargo period has elapsed, should this be the case.

---

<sup>14</sup> <http://www.invid-project.eu/publications>

## 3 Updated Quality Management Plan

This section reports the outcomes of the application of the InVID Quality Management Plan presented in D1.2 and provides updates, with respect to:

- innovation management;
- research activities;
- software development;
- project deliverables;
- overall progress and compliance with the timeplan of the project.

### 3.1 Innovation management

Following our Design Thinking methodology and through the knowledge acquired since the beginning of the project about the verification processes in journalism (an inspiration stage reinforced, among others, by the participation of InVID partner AFP in the First Draft News CrossCheck project on the French presidential election), we brought the idea of wrapping several useful tools into a browser plugin, conceptually designed as a kind of “Swiss army knife” to help journalists being more efficient, and to speed up the verification process by automating repetitive or cumbersome tasks.

The browser plugin was developed as a prototype, initially tested within the InVID consortium by AFP and DW journalists to get feedback. It was then rapidly released publicly to the verification community worldwide. We therefore reached the implementation stage (third stage of the Design Thinking methodology) by releasing on 3<sup>rd</sup> of July, in a First Draft News meeting, the plugin on the Chrome store.

The InVID browser plugin got very good feedback on social networks within the “verification community” (see indicative examples in Fig. 1). In the following weeks, we also released a Firefox version of the plugin as well as the plugin code under an MIT license on Github<sup>15</sup>. Through feedback and continuous testing, we updated the plugin again on 4<sup>th</sup> of August to enhance functionalities for end users on certain features like keyframes and metadata extraction, or Twitter advanced search. At the time of writing this deliverable (mid September 2017), the plugin has approximately 670 active users and the volume of videos handled through the plugin over the last 2.5 months for finding near duplicates with the help of the video fragmentation and image reverse search component of the tool, corresponds to more than 100 hrs of video data.

---

<sup>15</sup> <https://github.com/invideu/invid-verification-plugin>

Regarding the main platforms (verification application and discovery dashboard), we conducted several workshops or design and evaluation sessions to ensure that the usability of both prototypes meets the needs of the verification community.

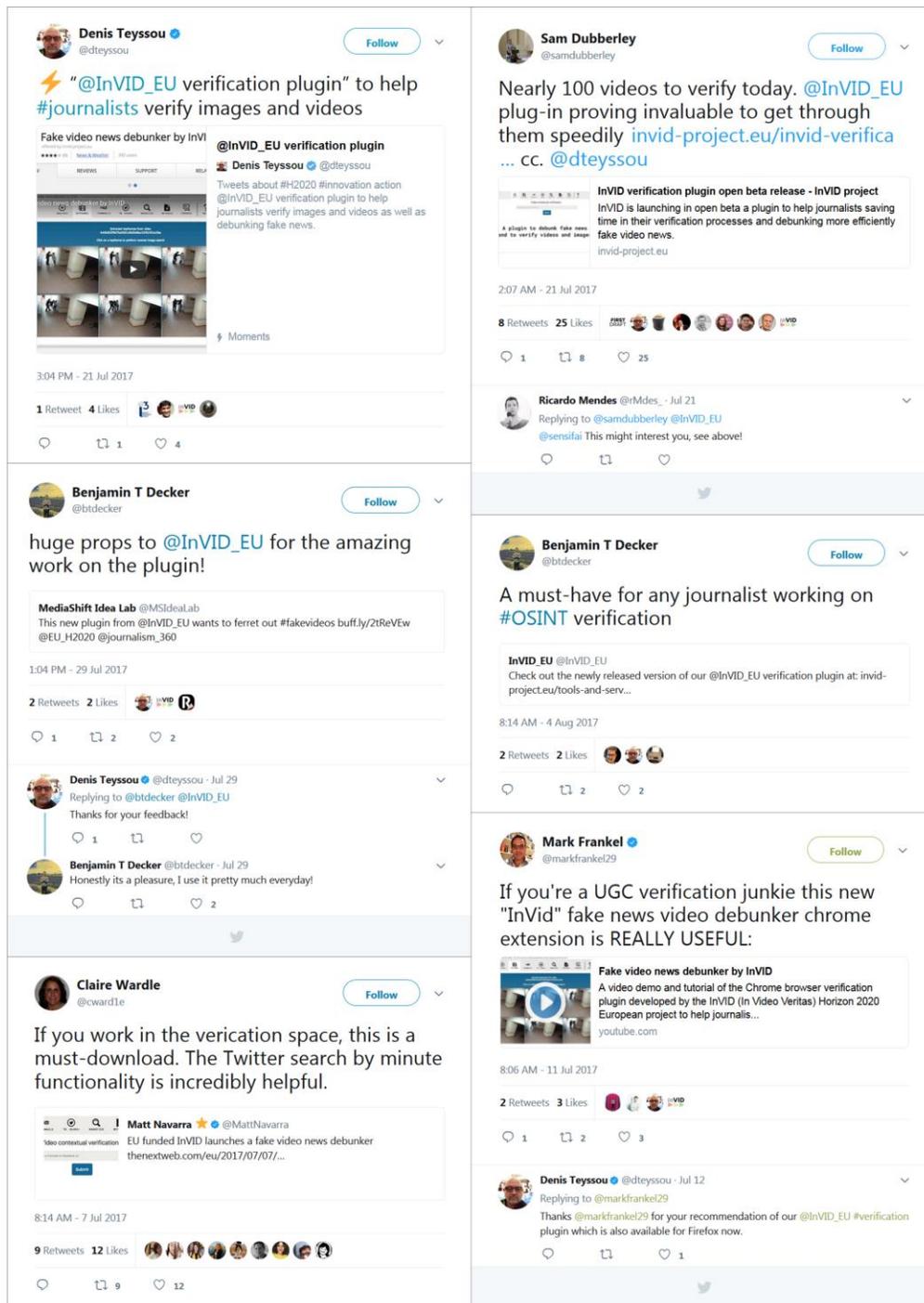


Figure 1: Received feedback on Twitter about the InVID browser plugin from members of the “verification community”

## 3.2 Research activities

The results of the conducted research activities within the project have been well appreciated by the relevant community. Manuscripts reporting on our developments for going beyond the current state-of-the-art on different research areas, such as machine learning, visual concept detection, near-duplicate video retrieval, ad-hoc video search, and contextual analysis, have been submitted and accepted for publication in 2 peer-reviewed journals and 16 international peer-reviewed conferences. The latter includes a common WP2-WP3 publication that describes the developed Browser Plugin (a.k.a. Verification Plugin) and has been accepted for publication at the 1<sup>st</sup> International Workshop on Multimedia Verification that will be held during the ACM Multimedia Conference, on Mountain View, CA, USA, in October 2017.

For facilitating access to these scientific results and increasing their exploitation and reuse (thus being aligned with the EC guidelines for Open Access to scientific data), we made them publicly available through the InVID Community on Zenodo<sup>16</sup> (while a full list of the project's publications is also available at: <http://www.invid-project.eu/publications/>). Moreover, the same online repository contains the created datasets for evaluating the efficiency of particular InVID technologies, and for fine-tuning them according to the analysis requirements of the project. As stated before, these datasets can be used for re-producing and validating the results reported in some InVID publications, following the principles of Open Research Data in Horizon 2020.

The same methodology will be followed by the project consortium concerning the research activities until the end of the project, unless currently unforeseen reasons introduce the need for a modification or adaptation of this plan.

## 3.3 Software development

The development of the InVID platform, application and individual software components was based on the agile development and validation cyclic approach presented in D1.1. The necessary individual technologies were built by WPs 2, 3 and 4 based on specified and prioritized industry requirements by WP6. Following up on this, WP5 and WP6 integrated these technologies into the InVID platform and applications. Finally, WP7 performed a number of testing cycles (four cycles so far) where each cycle evaluated and validated the latest released version of individual and integrated technologies, applications and software components.

---

<sup>16</sup> <https://zenodo.org/communities/invid-h2020>

Through the above mentioned procedure, the InVID project has built and evaluated a number of applications and web-based APIs (see Table 2 below), being fully aligned with the time-plan of the project (i.e. the project milestones) and the time-schedule for the development and validation of these InVID technologies (see Fig. 2).

**Table 2: List of developed and tested technologies (divided in Applications and APIs)**

Cycle	1	2	3	4
<b>Applications</b>				
InVID Multimodal Analytics Dashboard	Yes	Yes	Partially	Yes
Social Media Retrieval and Topic Detection	Yes	Yes	Yes	Yes
InVID Verification Application	Yes	–	Yes	Partially
Context Aggregation & Analysis	Yes	Yes	Yes	Yes
Video Fragmentation & Reverse Image Search	Yes	Yes	Yes	Yes
Logo Detection	Yes	Yes	Yes	Yes
Browser Plugin	–	–	Yes	Yes
Rights Management	–	–	Yes	Yes
InVID Mobile Application	–	–	–	Yes
<b>APIs</b>				
InVID Platform	Yes	Yes	Yes	Yes
Context Aggregation & Analysis Service	Yes	Yes	Yes	Yes
Video Fragmentation and Annotation Service	Yes	Yes	Yes	Yes
Logo Detection Service	Yes	Yes	Yes	Yes
Rights Management Service	Yes	Yes	Yes	Yes
Near Duplicate Detection Service	–	Yes	Yes	Yes

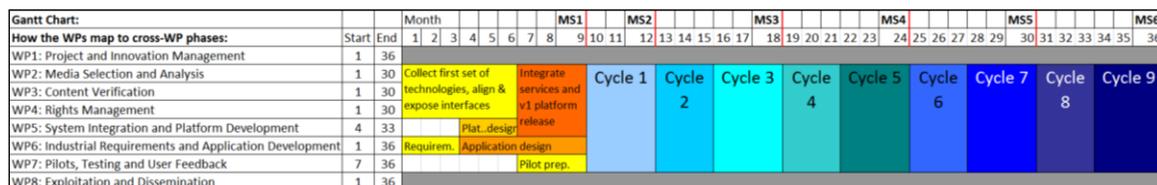


Figure 2: The nine agile development and validation cycles throughout the project

The same methodology will be followed by the project consortium concerning the software development until the end of the project, unless currently unforeseen reasons introduce the need for a modification or adaptation of this plan.

### 3.4 Project deliverables

For the preparation of the project deliverables, we followed the time-schedule described in Section 2.4 of D1.2. According to this time-schedule, a complete draft of the deliverable is submitted for Quality Assessment (QA) that aims to ensure its top quality and, at the very minimum, its compliance with the relevant contractual obligations set out in the Description of Action (DoA) document, and the documentation standards of the partners of the InVID consortium. As discussed in D1.2, in many cases we followed an approach that involves two Quality Assessors for each deliverable; in case of a technical deliverable, the Primary Quality Assessor had a technical background and the Secondary Quality Assessor a user/business background, while in case of a business-oriented deliverable, the reverse setup was used. After reviewing the deliverable, both Quality Assessors had to fill-in a Quality Assurance Review form that contains questions about the quality of the deliverable in terms of coverage, technical content, evaluation and innovation, and presentation, as presented in D1.2. Based on this feedback, the authors of the deliverable produced the final version of the document that was then submitted to the EC.

Following this plan, we prepared and submitted the first 17 project deliverables of the project on time (which were due in the first 18 months of the project's lifetime). The submitted deliverables and the responsible partners for their Quality Assessment are presented in Table 3 below. The same methodology will be followed by the project consortium concerning the preparation and submission of the remaining project deliverables until the end of the project, unless currently unforeseen reasons introduce the need for a modification or adaptation of this plan.

**Table 3: List of submitted deliverables**

ID	Title / File	QA	Due at
D8.1	<a href="#">Project website and promotional material</a>	AFP	M3
D1.1	<a href="#">Data Management Plan</a>	WLT	M4
D1.2	<a href="#">Quality and Knowledge Management Plan</a>	DW	M4
D6.1	<a href="#">InVID initial industrial requirements</a>	EXO & CERTH	M4
D2.1	<a href="#">Social media filtering and extraction</a>	APA-IT	M6
D5.1	<a href="#">Technology Roadmap</a>	Condat	M6
D8.2	<a href="#">Exploitation and dissemination plan and market analysis</a>	MODUL	M6

ID	Title / File	QA	Due at
D5.2	<a href="#">InVID Platform first version</a>	CERTH	M9
D3.1	<a href="#">Initial verification framework</a>	WLT & Condat	M12
D4.1	<a href="#">Overview of UGC Copyright Management industry practices</a>	DW & AFP	M12
D5.3	<a href="#">Multimodal Analytics Dashboard</a>	EXO & CERTH	M12
D6.2	<a href="#">Verification Application first release</a>	APA-IT & MODUL	M12
D2.2	<a href="#">Social media filtering and extraction pre-processing and annotation intermediate version</a>	EXO & AFP	M18
D4.2	<a href="#">Framework and Workflows for UGC Copyright Management</a>	APA-IT & WLT	M18
D5.4	<a href="#">InVID Platform intermediate version</a>	Condat & CERTH	M18
D7.1	<a href="#">Activities and outcome of the Pilots first report</a>	DW & CERTH	M18
D8.3	<a href="#">Exploitation and dissemination activities report and updated market analysis exploitation and dissemination plans</a>	UdL & MODUL	M18

### 3.5 Overall project progress and compliance with the timeplan

As stated in D1.2, the overall progress of the project according to the contractual obligations, and its compliance with the timeplan for fulfilling these obligations, is based on the list of project milestones (see Table 4). Based on coordinated activities by all WPs and partners of the project, and with the help of the established communication plan discussed in D1.2, the members of the Steering Board of the project (i.e. the Project Coordinator, the Innovation Manager and the Integration Manager) evaluated the compliance of the project's progress with this pre-defined time-plan.

The outcomes of this effort include successfully reaching the first three milestones of the project that relate to: a) the release and validation (through the first three testing cycles) of the InVID platform, b) the release, validation (through the 2<sup>nd</sup> and 3<sup>rd</sup> testing cycles) and partial demonstration of the InVID Verification Application, c) the release of the UGC management system that integrates the individual analysis components of the InVID system, and d) the delivery of 17 project deliverables. At this time of the project's life (M21) we are working on meeting the goals of the 4<sup>th</sup> milestone of the project, by

assessing and validating the latest versions of the InVID platform, applications and individual components through the 4<sup>th</sup> testing cycle of the project.

**Table 4: List of project's milestones**

<b>MS#</b>	<b>MS title</b>	<b>Due at</b>	<b>Means of verification</b>
MS1	First platform release	M9 <b>Done</b>	The first release of the InVID platform is online and functional. Delivery of D1.1, D1.2, D2.1, D5.1, D5.2, D6.1, D8.1, D8.2.
MS2	First platform validation and initial release of verification application	M12 <b>Done</b>	The first release of the InVID platform is validated by completion of the first development and validation cycle (Cycle 1). The first prototype of the InVID Verification Application is functional. Delivery of D1.3, D3.1, D4.1, D5.3, D6.2.
MS3	Updated platform and verification application validation and demonstration in relevant environment	M18 <b>Done</b>	Completion of the third development and validation cycle (Cycle 3). The second version of the InVID Verification Application and the first versions of the UGC management systems integrating it are functional, and the Verification Application is validated and partially demonstrated. Delivery of D2.2, D4.2, D5.4, D7.1, D8.3.
MS4	InVID updated application and system prototypes demonstrated in operational environment	M24	Completion of the fifth development and validation cycle (Cycle 5). The updated InVID Verification Application and UGC management systems integrating it are functional and demonstrated. Delivery of D1.4, D3.2, D4.3, D6.3.
MS5	InVID fully-fledged application and system prototypes demonstrated in operational	M30	Completion of the seventh development and validation cycle (Cycle 7). All verification functionalities are integrated, and the complete InVID Verification Application and UGC management systems integrating it are

MS#	MS title	Due at	Means of verification
	environment		demonstrated in operational environment. Delivery of D2.3, D7.2.
MS6	InVID final application and systems complete and qualified	M36	Completion of the final development and validation cycle (Cycle 9). The complete, optimized InVID Verification Application and UGC management systems integrating it are qualified. Delivery of D1.5, D3.3, D4.4, D5.5, D6.4, D7.3, D8.4.

Moreover, as reported in D8.3, the InVID project was highly productive in terms of dissemination. It raised international awareness through its participation in the First Draft News partner network and a number (currently more than 35) of events involving experts from the EC Community, the research and academic sectors, industry and the general public. Through these coordinated efforts by the members of the project consortium, and the efficient use of different communication instruments of the project (e.g. the project website, the InVID Twitter account, the InVID Slideshare channel and the InVID community on Zenodo) a number of pre-determined key performance indicators (KPI) regarding the dissemination of the project activities and the created scientific data have been surpassed. As presented in Section 4.3 of D8.3, most of the defined success metrics for evaluating the effectiveness of our dissemination activities during M7-M18 of the project's life have been achieved or even exceeded. The only goals that were initially not met fully were related to the activity of the project website, and more specifically to the number of page views and visits per month.

Nevertheless, as shown in Table 5 below, the release of the Browser Plugin and its dissemination through the project website and the other communication means of the project (such as the InVID account on Twitter<sup>17</sup>) that also point to the central webpage about the plugin<sup>18</sup> led to a remarkable increase in the number of page views and visits over the last three months of the project. Based on this increased access on the project website, the KPI related to the visits per month is now surpassed by 12%, while the number of page views per month is now closer to the defined goal, being increased by 40%. Finally, the other three KPIs regarding the average visit duration, the percentage of

<sup>17</sup> [https://twitter.com/InVID\\_EU](https://twitter.com/InVID_EU)

<sup>18</sup> <http://www.invid-project.eu/tools-and-services/invid-verification-plugin/>

returning visitors and the number of countries from where the website has been visited, are fairly constant over the determined success metrics.

**Table 5: KPI about the project website versus real figures**

Measured Indicator	KPI from D8.2	Real Figures (early Sept. 2017)
<b>Project Website</b>		
Page views per month	>1000	~742
Visits per month	>250	~280
Average visit duration	>1 min 30 sec	~2min 30sec
Returning visitors	>30%	~35%
Countries from where the website is visited	>50	108

## 4 Updated Knowledge Management Plan

The InVID Knowledge Management Plan presented in Section 3.1 of D1.2 is still fully applicable, a fact that highlights the suitability and long-term efficiency of the established plan in the beginning of the project. Given this fact, the following subsections update and extend (per work package) the set of the Intellectual Property (IP) assets discussed in D1.2, by reporting: i) the date when the IP asset was available (including also data about different versions of it), and ii) the sharing status and any future plan for sharing it. Moreover, newly-introduced IP assets, such as the Browser Plugin, are introduced.

### 4.1 WP2 IP assets

IP asset	<b>InVID_IP_WP2_1_TopicDetectionComponent</b>
Type	Foreground
Description	This IP asset will extract from a set or stream of documents (e.g. news articles, tweets) a list of detected "topics", where topics refers to clusters of documents which can be said with some confidence to jointly relate to a specific news event or story.
Ownership	MODUL Technology
Protection	This is proprietary software licensed by MODUL Technology.
Access rights	Access rights will be granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	First deployed in the webLyzard platform at M6 (June 2016), ongoing development until M30 (June 2018).
Sharing & sharing status	A pre-configured version of this asset will be made available for public access (Story API) in co-operation with webLyzard during the 3 <sup>rd</sup> year of InVID, with an initial limited release to InVID partners

IP asset	<b>InVID_IP_WP2_2_VideoFragmentation&amp;AnnotationComponent</b>
Type	Foreground
Description	This component is a web service with a REST API, that performs temporal decomposition of a video into three different levels of granularity; scenes (i.e. semantically and temporally coherent segments that correspond to the story-telling parts of the video), shots (i.e. sequences of frames captured uninterruptedly by a single camera) and sub-shots (i.e. sub-parts of a shot with visually discrete content; useful when analyzing single-shot videos). Following, it identifies the semantics of the video at the shot- and sub-shot-level by detecting a number of high level visual concepts after analyzing one representative keyframe per video fragment. In this way, this component produces a fragment-level concept-based annotation for each analysed video. This component has been integrated in the InVID Verification Application and the InVID Multimedia Analytics Dashboard.
Ownership	CERTH
Protection	This is proprietary software that uses technologies developed by CERTH.
Access rights	Access rights have been granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	The first version was made available in M11 (Nov. 2016). The second version was made available in M14 (Feb. 2017). The third version was made available in M17 (May 2017). The fourth version was made available in M19 (Jul. 2017). The final version will be available by M30 (June 2018) and will be reported in D2.3.
Sharing & sharing status	Several versions have already been shared with the consortium members in M11, M14, M17 and M19 (see above) for testing and integration purposes. Any updated version based on the feedback from the future test cycles of the project will be shared with the project partners.

IP asset	<b>InVID_IP_WP2_3_VideoFragmentation&amp;ReverseImageSearch Component</b>
Type	Foreground
Description	This component is a web application that allows the user to extract a set of representative keyframes from a video, and to use these keyframes for performing reverse image searches. To submit a video for analysis, the user can either provide its URL or upload a local copy of it from his/her machine <sup>19</sup> . After submitting and analysing the video, the user is shown the collection of extracted keyframes and is able to perform reverse image search via the Google Image Search engine, by simply clicking on each one of them. This component can be used as a standalone tool <sup>20</sup> , while it has been integrated in the Browser Plugin.
Ownership	CERTH
Protection	This is proprietary software that uses technologies developed by CERTH.
Access rights	Access rights have been granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	The first version was made available in M11 (Nov. 2016). The second version was made available in M14 (Feb. 2017). The third version was made available in M17 (May 2017). The fourth version was made available in M19 (Jul. 2017). The final version will be available by M30 (Jun. 2018) and will be reported in D2.3.
Sharing & sharing status	Several versions have been already shared with the consortium members in M11, M14, M17 and M19 (see above). The latest release (fourth version) was made also publicly available in M19.

<sup>19</sup> All video rights remain with the uploader, who is assumed to have the right to submit the video to this service for analysis.

<sup>20</sup> Accessible at: [http://multimedia3.iti.gr/videofragmentation\\_v5/service/start.html](http://multimedia3.iti.gr/videofragmentation_v5/service/start.html)

## 4.2 WP3 IP assets

IP asset	<b>InVID_IP_WP3_1_VideoForensicsComponent</b>
Type	Foreground
Description	This component aims to help assessing if a video has been tampered or not. A first version of this component was built on existing expertise and technologies (TUNGSTEN) for image forensics detection, after extending these filters for tampering detection in videos. Furthermore, new video-oriented filters have been developed, related to multiple quantisation detection (e.g. Q2 filter), focus analysis (e.g. Focus filters), DCT block characterisation (e.g. Q4 and Q4 order filters), illumination analysis (e.g. Spectrum Locus filter), MPEG-4 conformance (e.g. Cobalt filters), temporal behaviour and recovering and lighting enhancement (e.g. HELIUM3 filter family). The output of these filters is currently used to train a system that aims to provide indication of tampering in a video, in a more automatic way.
Ownership	EXO MAKINA
Protection	This is proprietary software that uses technologies developed by EXO MAKINA.
Access rights	Access rights have been granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	A first version containing the image-compatible filters was released in M12 (Dec. 2016) and reported in D3.1. The final video-compatible version will be released in M36 (Dec. 2018) and reported in D3.3, with intermediate updates following every test cycle.
Sharing & sharing status	An online instance of this component has been shared internally with the project partners for testing and integration purposes. An open source contribution concerning specific filters of this component to the FFMPEG project is under discussion.

IP asset	<b>InVID_IP_WP3_2_VideoLogoDetectionComponent</b>
Type	Foreground
Description	This component accepts videos from various social media sources and video sharing platforms, and identifies logos based on a list of known logo templates. It then presents the user with information about detected logos, such as the name of the associated organization/company, and the corresponding Wikipedia link. It also provides a tool for submitting new logo templates for future recognition. A first version of the module has been developed and presented in D3.1. An improved version based on feedback from the Test Cycles and further research and development from CERTH will be presented in D3.2.
Ownership	CERTH
Protection	This is proprietary software that uses technologies developed by CERTH.
Access rights	Access rights have been granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	The first version was released in M12 (Dec. 2016) and reported in D3.1. The final version will be released in M36 (Dec. 2018) and reported in D3.3, with intermediate updates following every test cycle.
Sharing & sharing status	A restricted demo service has been made publicly available <sup>21</sup> .

IP asset	<b>InVID_IP_WP3_3_NearDuplicateDetectionComponent</b>
Type	Foreground

<sup>21</sup> <http://logos.iti.gr/>

IP asset	<b>InVID_IP_WP3_3_NearDuplicateDetectionComponent</b>
Description	This component accepts video queries and searches within a dataset of indexed videos for near-duplicates. It also provides API calls for submitting new videos to the dataset and removing videos from it. A first version of the module has been developed and presented in D3.1. An improved version based on feedback from the test cycles and further research and development from CERTH will be presented in D3.2
Ownership	CERTH
Protection	This is proprietary software that uses technologies developed by CERTH.
Access rights	Access rights have been granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	The first version was released in M12 (Dec. 2016) and reported in D3.1. The final version will be released in M36 (Dec. 2018) and reported in D3.3, with intermediate updates following every test cycle.
Sharing & sharing status	Currently, only API service calls are given, and sharing has taken place only within the consortium. A restricted demo service may be made publicly available in the future.

IP asset	<b>InVID_IP_WP3_4_ContextualVerificationComponent</b>
Type	Foreground
Description	This component accepts videos published on YouTube or Facebook, and provides a list of verification clues aiming to assist investigators in verifying its veracity. A first version of the module has been developed and presented in D3.1. An improved version based on feedback from the Test Cycles and further developments from CERTH will be presented in D3.2.
Ownership	CERTH

IP asset	<b>InVID_IP_WP3_4_ContextualVerificationComponent</b>
Protection	This is proprietary software that uses technologies developed by CERTH.
Access rights	Access rights have been granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	The first version was released in M12 (Dec. 2016) and reported in D3.1. The final version will be released in M36 (Dec. 2018) and reported in D3.3, with intermediate updates following every test cycle.
Sharing & sharing status	A restricted demo service has been made publicly available <sup>22</sup> .

IP asset	<b>InVID_IP_WP3_5_LocationDetectionComponent</b>
Type	Foreground
Description	This will be a web service which is trained to extract location information from social media content (e.g. title and description).
Ownership	MODUL
Protection	This service will extend MODUL's license-protected RECOGNISE tool for Named Entity Recognition and Linking.
Access rights	The service will be made available under the same licensing conditions as RECOGNISE, which includes access rights to all InVID partners during the project and third party licensing possible after the project.
Dates of IP	First internal deployment and testing at M12 (Dec. 2016), planned

<sup>22</sup> <http://caa.iti.gr/>

IP asset	<b>InVID_IP_WP3_5_LocationDetectionComponent</b>
asset availability	production-level deployment in the webLyzard platform by M24 (Dec. 2017), planned limited-release API-based service for third parties (beginning with selected InVID partners) in the third year of InVID.
Sharing & sharing status	A restricted demo service may be made publicly available. After production level release and successful deployment on the webLyzard platform, we will consider third party access via API, beginning with InVID partners.

### 4.3 WP4 IP assets

IP asset	<b>InVID_IP_WP4_1_CopyrightOntology</b>
Type	Foreground
Description	This will be an update of the existing Copyright Ontology, featuring concepts and relationships that improve its expressiveness in the context of UGC and the News Industry.
Ownership	UdL
Protection	Creative Commons, open for non-commercial uses.
Access rights	Non-commercial uses will be granted if attribution is performed. Other uses might be granted to InVID partners for the needs of the project. Commercial uses beyond what has been previously specified will be agreed in separate business agreements.
Dates of IP asset availability	Though the original version of the ontology has been available since the beginning of the project, the first operative version tailored to InVID will be available in M24 (Dec. 2017) together with D4.3. The final version will be available with D4.4 in M36 (Dec. 2017).
Sharing & sharing	The basic version has been available under a Creative Commons Non-Commercial and Share-Alike (CC-NC-SA) license since the beginning of

IP asset	<b>InVID_IP_WP4_1_CopyrightOntology</b>
status	the project. The first InVID version of the ontology will be available under the same terms in M24 and the final version, also licensed as CC-NC-SA, in M36.

IP asset	<b>InVID_IP_WP4_2_CopyrightManagementModule</b>
Type	Foreground
Description	This module implements the requirements described in D6.1 related to rights management, Section 4.7. These requirements are specified in D4.2. The module will generate and use the dataset "InVID_Data_WP4_1_UGCRegisteredProviders" as described in D1.1.
Ownership	UdL
Protection	This is proprietary software of UdL.
Access rights	Access rights will be granted to the InVID partners for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	A preliminary and partial implementation of the module has been available since M18 (June 2017) as part of D4.2. A more complete and functional implementation will be available in M24 (Dec. 2017) together with D4.3. The final version will be available with D4.4 in M36 (Dec. 2018).
Sharing & sharing status	A restricted demo service has been publicly available since M18. This includes the API, which is available internally to the rest of the partners, and a Web client, which is publicly available. However, some functionalities of the API (and thus of the Web client that has been built on top of it) are password protected. The consortium members have had access to these passwords since M18.

#### 4.4 WP5 IP assets

IP asset	<b>InVID_IP_WP5_1_ApplicationProgrammingInterface</b>
Type	Foreground
Description	This IP asset will be used to: i) manage the addition, modification, annotation, and deletion of data elements stored within the InVID knowledge repository, comprising both unstructured data in the form of text documents (e.g. postings that accompany user-generated videos, transcripts, related news articles) and structured data in the form of numeric indicators, ii) query the central knowledge repository along multiple metadata dimensions, and iii) visualize aggregated content and metadata patterns.
Ownership	webLyzard technology
Protection	This is proprietary software licensed to webLyzard technology.
Access rights	Access rights will be granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	First made available in M9 (Sept. 2016) and reported in D5.2; the final version will be available in M36 (Dec. 2018).
Sharing & sharing status	Already shared with consortium members in M9.

IP asset	<b>InVID_IP_WP5_2_MultimodalAnalyticsDashboard</b>
Type	Foreground
Description	This IP asset represents an extended version of the existing user interface of the webLyzard platform, adding a range of interactive features and the

IP asset	<b>InVID_IP_WP5_2_MultimodalAnalyticsDashboard</b>
	ability to display image and video content and use thumbnails to represent related stories and content clusters. Recent InVID additions and extensions include: i) increased data granularity supporting hourly intervals; ii) moving average computation in 7/30/60 day intervals; and iii) multimedia support including thumbnail display and video playback for multiple social media sources; iv) a PDF report generator to create automated summaries of analytic results, increasing the utility of the platform and widening its target group beyond actual users of the dashboard.
Ownership	webLyzard technology
Protection	This is proprietary software licensed by webLyzard technology.
Access rights	Access rights will be granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	First made available in M12 (Dec. 2016) and reported in D5.3; the final version will be available in M24 (Dec. 2017).
Sharing & sharing status	Already shared with consortium members in M12.

IP asset	<b>InVID_IP_WP5_3_VisualizationComponents</b>
Type	Foreground
Description	This IP asset develops new (and extends existing) interactive visualization tools that will be made available in two different formats: fully integrated into the Multimodal Analytics Dashboard (IP_WP5_IP2; all modules), and as embeddable widgets to be integrated into InVID applications (selected modules). Recent InVID additions and extensions include: i) <i>Story View</i> to

IP asset	<b>InVID_IP_WP5_3_VisualizationComponents</b>
	represent document clusters - including a lead article or posting, and related content from chosen sources; ii) Story Flow and Story Cluster as visual representations of the temporal distribution and structural composition of document clusters – three variations of this component were developed (episodes, burst, streamgraph) and assessed as part of the 4 <sup>th</sup> test cycle.
Ownership	webLyzard technology
Protection	This is proprietary software licensed by webLyzard technology.
Access rights	Access rights will be granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	First made available in M12 (Dec. 2016); the final version will be available in M24 (Dec. 2017).
Sharing & sharing status	Already shared with consortium members in M12.

#### 4.5 WP6 IP assets

IP asset	<b>InVID_IP_WP6_1_Verification_App</b>
Type	Foreground
Description	This component uses the different underlying InVID components to perform the verification of a video.
Ownership	Condat
Protection	This is proprietary software that uses technologies developed by other

IP asset	<b>InVID_IP_WP6_1_Verification_App</b>
	InVID partners.
Access rights	Access rights have been granted to the InVID participants for the needs of the project. Rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral Business Agreements.
Dates of IP asset availability	The first version was released in M12 (Dec. 2016) and reported in D6.1. The final version will be released in M36 (Dec. 2018) and reported in D6.4, with intermediate updates following every test cycle.
Sharing & sharing status	A first version of the Verification Application has been shared with the project partners for testing and evaluation purposes <sup>23</sup> . A password-protected version of the Verification Application will be made publicly available in early 2018.

IP asset	<b>InVID_IP_WP6_2_VerificationAppExpertise</b>
Type	Background
Description	Deutsche Welle brings into the project its know-how in gathering real-life journalistic requirements and turning them into functionalities of the InVID platform and its applications (thereby assisting technology / module / application development). Special attention is paid to the development of intuitive graphical user interfaces, which are considered crucial for the commercial exploitation of the InVID products and services. Deutsche Welle falls back in particular on its expertise gained in the projects REVEAL, an EC-funded project dealing with the verification of UGC, and Truly Media, a project partially financed by Google's Digital News Initiative that provides a UGC verification and collaboration platform.

---

<sup>23</sup> For getting access to this version of the Verification Application, please contact Rolf Fricke (Rolf.Fricke@condat.de) or Jan Thomsen (Jan.Thomsen@condat.de).

IP asset	<b>InVID_IP_WP6_2_VerificationAppExpertise</b>
Ownership	Deutsche Welle
Protection	Expertise of Deutsche Welle staff and feedback brought into InVID developments (especially via validation and testing cycles).
Access rights	Deutsche Welle shares its know-how with the project partners in the validation activities and the development of the InVID platform, its single components and the Verification Application.
Dates of IP asset availability	Not applicable.
Sharing & sharing status	See above (access rights).

#### 4.6 WP7 IP assets

IP asset	<b>InVID_IP_WP7_1_UGCMobileApp</b>
Type	Foreground
Description	This IP asset has been created to enable users to upload videos for the showcase with regional newspapers. It uses the user base managed by the regional newspaper. The user is asked to accept the usage terms and to allow additional data to be uploaded (geo-data, device-information etc). In a later version back-channels to be used by the newspapers will also be integrated.
Ownership	APA-IT
Protection	This is proprietary software that is going to use already existing technologies of APA-IT's Mobile Publishing Suite.

IP asset	<b>InVID_IP_WP7_1_UGCMobileApp</b>
Access rights	Access rights will be granted to the local newspapers for the pilots at no costs. If needed, access rights for other InVID participants will be granted although adaptations – e.g. interfaces to specific user-databases – as well as rights pertaining to joint exploitation activities will be agreed upon between the partners in separate bilateral business agreements.
Dates of IP asset availability	First version was made available in M20 (Aug. 2016); the final version will be available in M36 (Dec. 2018).
Sharing & sharing status	The UGC Mobile Application has been shared with the consortium members in M21 (Sep. 2017) and will be shared with external stakeholders in M23 (Nov. 2017).

IP asset	<b>InVID_IP_WP7_2_BrowserPlugin (a.k.a. Verification Plugin)</b>
Type	Foreground
Description	The IP asset has been created to help journalists in their verification process by wrapping several tools in one free browser extension. The aim is to bring tools to the “verification community”. The code of the browser extension has been released under an MIT license. The plugin is part of a freemium strategy to raise awareness about the project, to engage with end-users and to gather feedback from them in order to enhance and improve forthcoming services.
Ownership	AFP (as part of the AFP pilot)
Protection	The plugin code is open source and has been released under an MIT license. This does not include the underlying services available through API calls.
Access rights	Access rights are granted to anybody. So far, journalists, human rights organisations, non-governmental organizations and media educators are

IP asset	<b>InVID_IP_WP7_2_BrowserPlugin (a.k.a. Verification Plugin)</b>
	testing and using it.
Dates of IP asset availability	First made available in M18 (June 2017); the final version will be available in M36 (Dec. 2018).
Sharing & sharing status	See above (access rights).

## 5 Summary

In this deliverable, the updated Data, Quality and Knowledge Management Plan of InVID was reported. In terms of Data Management, datasets included in the first release of the Data Management Plan (reported in D1.1) and are no longer relevant to the activities of the project have been reported, and newly created or updated datasets were presented (Section 2). In terms of Quality Management, the outcomes of the foreseen procedures related to the innovation management, the research activities, the software development, the preparation of the project deliverables and the compliance of the project with the determined timeplan were discussed, and updates of the first release of the Quality Management Plan (reported in D1.2) were provided wherever necessary (Section 3). Finally, in terms of Knowledge Management, the first release of the Knowledge Management Plan (reported in D1.2) is still fully applicable, a fact that highlights the suitability and efficiency of the established plan in the beginning of project's life. Based on this, the current document updated and enriched the part of this plan that relates to the Intellectual Property assets of the InVID project (Section 4).