

ファイル共有ソフトの利用に関する調査
～クローリング調査～

報 告 書

2011年2月

社団法人コンピュータソフトウェア著作権協会

目 次

I. 調査概要	1
1. 調査の前提	1
(1) 基礎情報	1
2. データの抽出	1
(1) フィルタリング	1
(2) 権利の対象性の算出方法	2
II. 調査結果	3
1. Winny	3
(1) コンテンツの流通状況	3
(2) 権利の対象性	5
(3) 検出ノードの国・地域	8
1. Share	9
(1) コンテンツの流通状況	9
(2) 権利の対象性	11
(3) 検出ノードの国・地域	14
1. PerfectDark	15
(1) コンテンツの流通状況	15
(2) 権利の対象性	17
(3) 検出ノードの国・地域	20

I. 調査概要

1. 調査の前提

調査は2010年12月17日 17:00から2010年12月18日 17:00の24時間、以下のP2Pネットワークに対応した手法を用いてネットワークを巡回(クローリング)し、実際にネットワーク上を流通している情報を自動収集、分析する形で実施した。

Winny、Share、PerfectDarkそれぞれのプロトコルを利用したクローラを用いて、特にキーワードを設定することなくWinnyネットワーク上に流通するキー情報(ノード情報、ファイル情報)の自動収集を行った。複数のクローラを用いる事で、24時間でほぼネットワークの全域をクローリングできる性能を確保している。

(1) 基礎情報

	Winny	Share	PerfectDark
利用したソフトウェア	P2P FINDER(Winny) 2009年9月Version	P2P FINDER(Share) 2009年9月Version	P2P FINDER(PerfectDark) 2010年4月Version
設定情報	総スレッド数 1,200	総スレッド数 3,000	総スレッド数 300

2. データの抽出

(1) フィルタリング

ファイル共有ソフトネットワーク上で、権利者に無許諾で送信可能な状態におかれ、流通しているファイルの調査を行った。

調査を行うにあたり、総取得件数からノード(IPとポート)およびファイル名が同一なデータを取り除いた後、(調査対象データ)、2万件をランダムに抽出した。調査対象データに対する抽出データ(2万件)の割合はそれぞれWinny:0.023%、Share 0.327%、PerfectDark 2.783%となった。調査対象データから、アダルト系キーワード、共通除外キーワードを含むデータを除外し、データを目視にて確認し、各ファイルについて推定されるジャンル、権利の対象および許諾の有無について調査した。

	工程
①総取得件数	クローラによりIP、ポート、ファイル名、時間を取得
②重複件数の削除	①で取得したデータのうち、IP、ポート番号、ファイル名が重複したデータを削除(PerfectDarkについてはファイルネーム、化身ID)
③間引き後件数	②で取得したデータを20,000件になるようにランダムに抽出
④アダルトキーワード除去	③で抽出したデータのうちファイル名にアダルトコンテンツと想定されるキーワードがあるデータを除外
⑤共通除外キーワード除去	④のデータのファイル名に共通除外キーワードがあるデータを除外
⑥合法ファイル抽出	④のデータのファイル名に「合法」のキーワードがあるデータを抽出
⑦キーワード抽出	⑤のデータを各ジャンルのキーワードで抽出

(2) 権利の対象性の算出方法

2.(1)で抽出したデータを目視にて以下のジャンルに分類を行った。

- ・ 著作物と推定されるもの
- ・ アダルト
- ・ 同人
- ・ 不明ファイル
- ・ 危険ファイル
- ・ 合法ファイル

分類の際には、抽出データ（2万件）の目視確認を行い、2.(1)のキーワードフィルタリングで漏れたコンテンツのジャンルの修正も行った。

II. 調査結果

1. Winny

本調査でIPアドレスとポート番号の一意的組み合わせをノードの量として算出したところ、全数として約6万ノードと推定される。

(1) コンテンツの流通状況

流通コンテンツのうち、およそ半数のコンテンツが著作物と推定される。

著作物と推定されるもの	9,475	音楽	1,676		
		映像	3,866	映画	45
				映画(海外)	74
				映像	851
				映像(海外)	152
				アニメ	2,506
				PV	238
				プログラム	337
				ゲーム	217
		書籍	2,483	書籍	513
コミック	1,970				
その他	1,113				
アダルト	6,101				
同人	3,311				
不明ファイル	595				
危険ファイル	31				
情報	487				
計	20,000				

※「著作物と推定されるもの」とは本調査で権利の所在が推定できるもの

※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

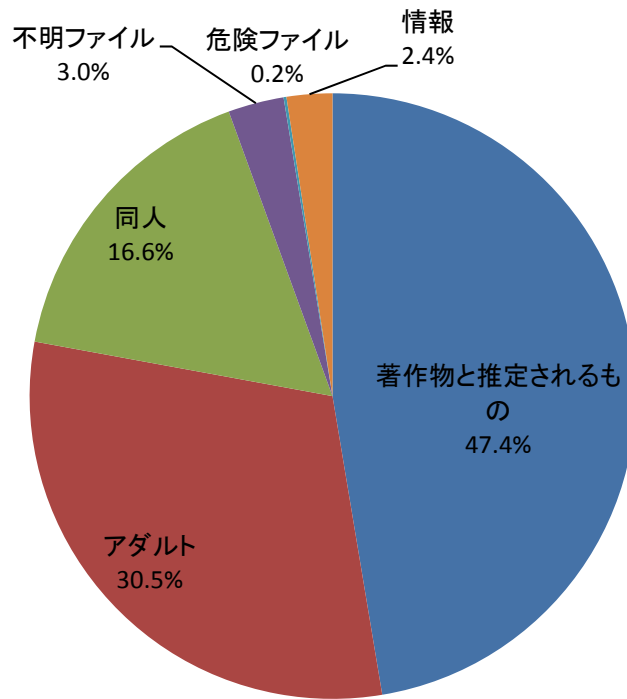
※「不明ファイル」とはタイトルからコンテンツの内容が判別できないもの

※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

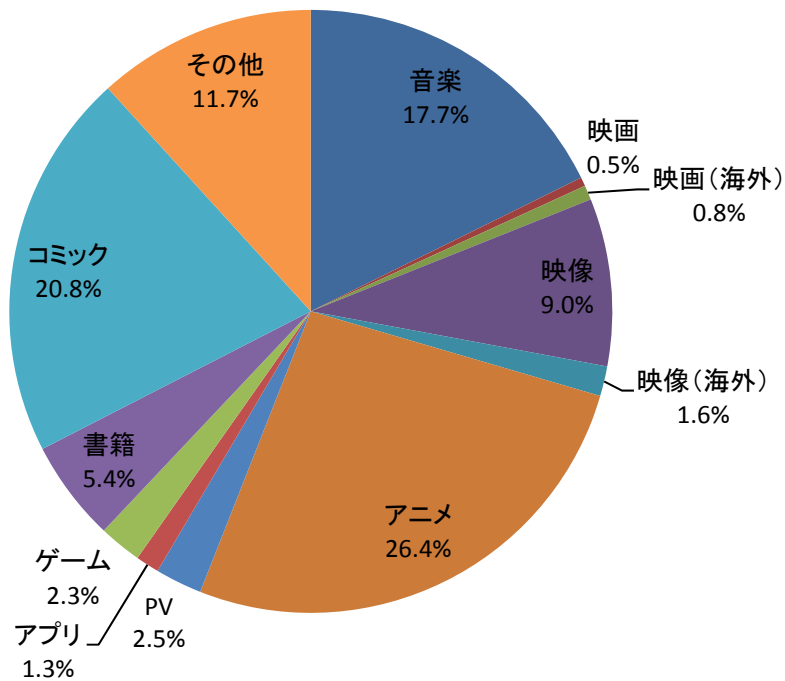
※「情報」とはウイルス感染などで流出した個人・組織等の情報だと推定されるもの

※「その他」とはコンテンツの分類が音楽、映像関連、プログラム、書籍関連に含まれないもの

Winny コンテンツ流通状況



Winny 著作権と推定されるコンテンツの内訳

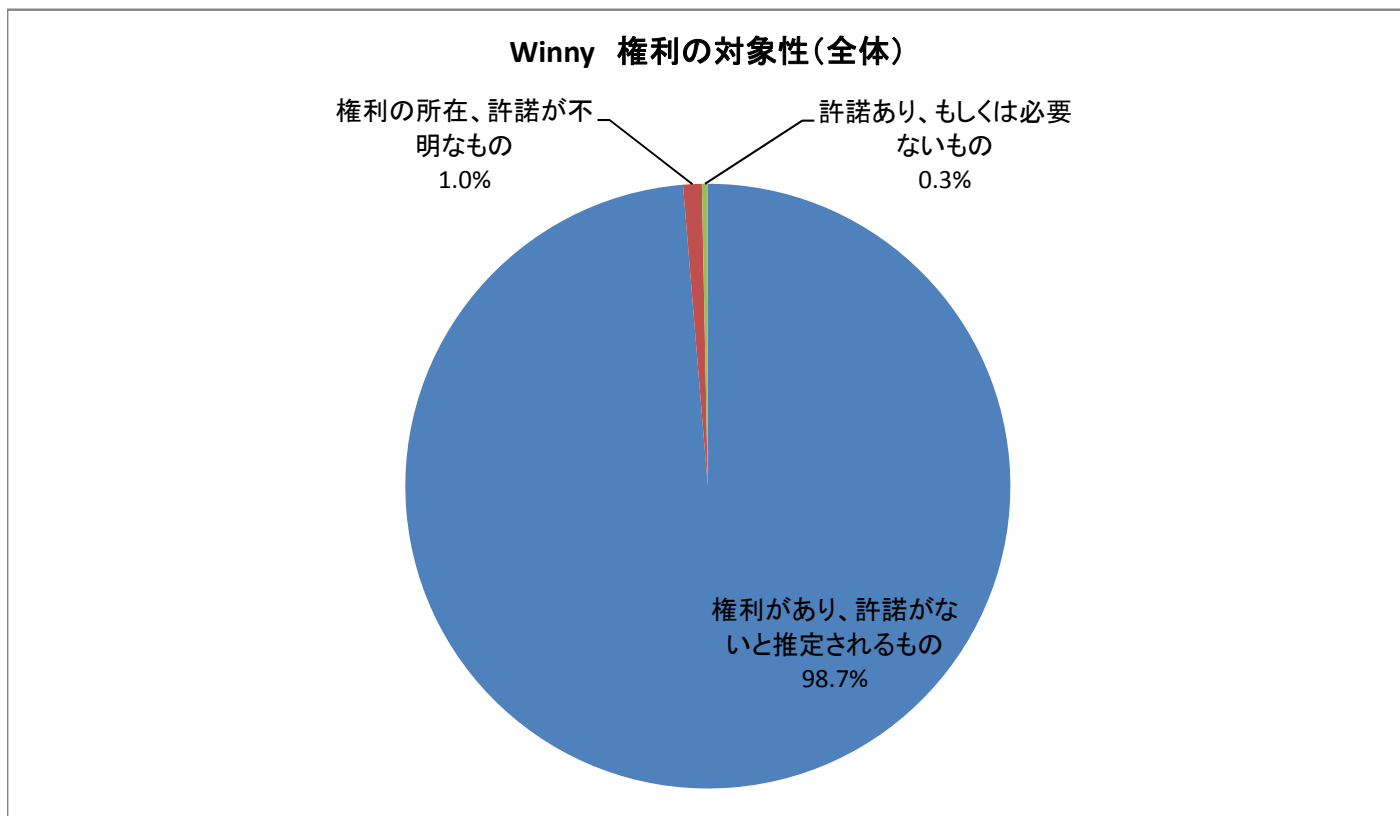


(2) 権利の対象性

① 全体

著作物と推定されるもののうち、98.7%に権利があり、かつ許諾がないものと推定される。

権利があり、許諾がないと推定されるもの	9,351
権利の所在、許諾が不明なもの	94
許諾あり、もしくは必要ないもの	30

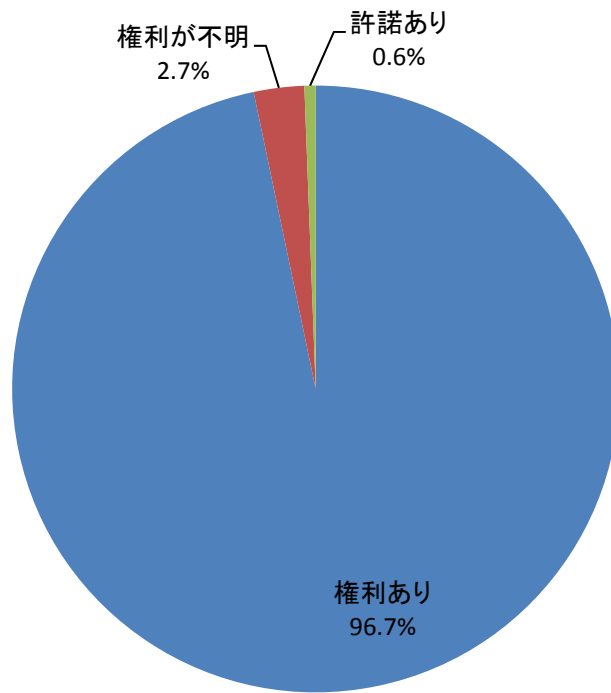


② ジャンルごと

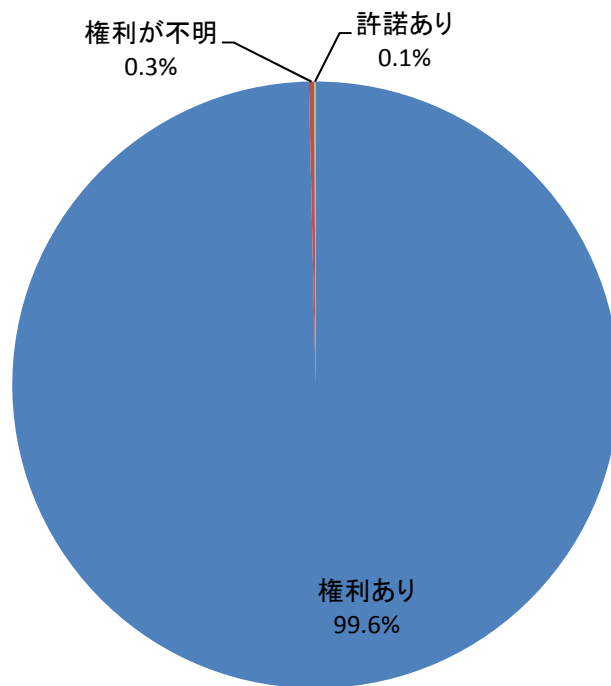
権利の対象性について、コンテンツジャンルごとに集計した。全ジャンルで90%以上が権利の対象であり、無許諾で送信されていると推定されるものであった。

	権利あり	権利が不明	許諾あり
音楽	1,621	45	10
映像	3,852	11	3
プログラム	315	8	14
書籍	2,480	2	1
その他	1,083	28	2

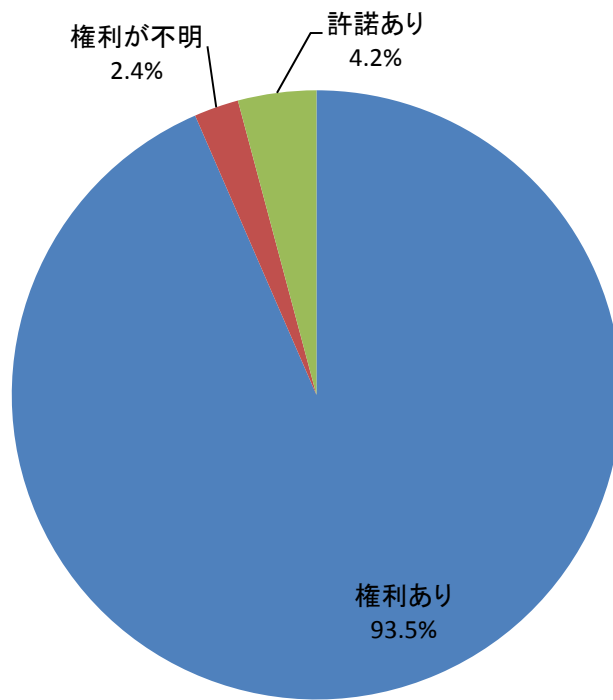
Winny 権利の対象性(音楽)



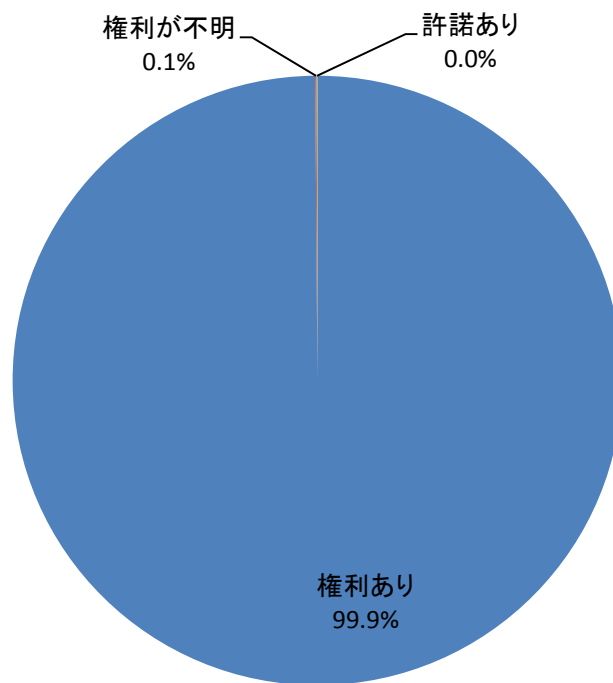
Winny 権利の対象性(映像)



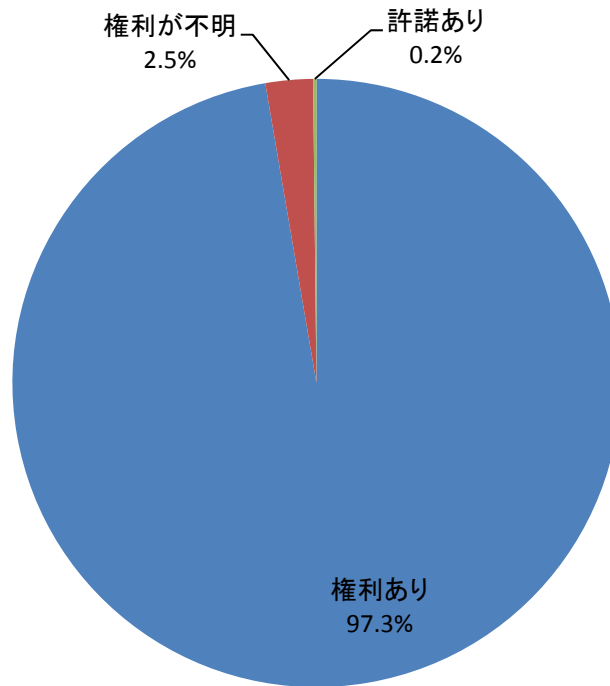
Winny 権利の対象性(プログラム)



Winny 権利の対象性(書籍)

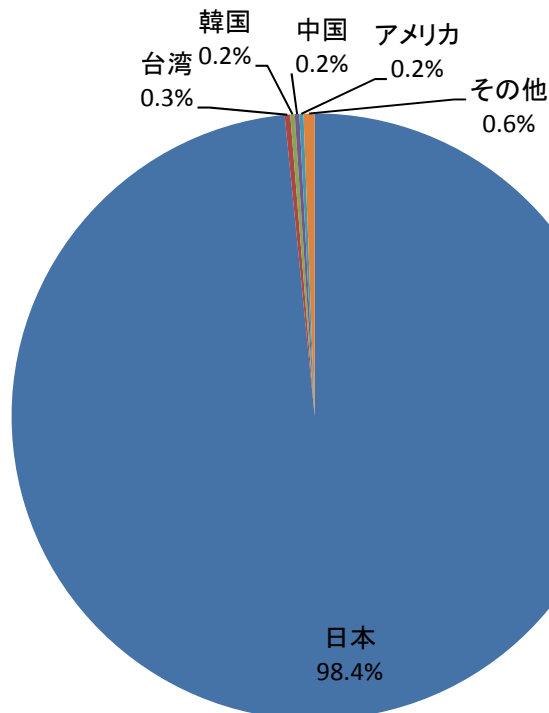


Winyy 権利の対象性(その他)



(3) 検出ノードの国・地域

Winyy 検出ノード国・地域分布



2. Share

本調査でIPアドレスとポート番号の一意的な組み合わせをノードの量として算出したところ、全数として約13万ノードと推定される。

(1) コンテンツの流通状況

流通コンテンツのうち、およそ半数のコンテンツが著作物と推定される。

著作物と推定されるもの	10,714	音楽	847		
		映像	6,833	映画	208
				映画(海外)	791
				映像	1,403
				映像(海外)	520
				アニメ	3,680
				PV	231
		プログラム	463	アプリ	207
				ゲーム	256
		書籍	2,155	書籍	333
コミック	1,822				
その他	416				
アダルト	6,039				
同人	2,114				
不明ファイル	1,041				
危険ファイル	21				
情報	71				
計	20,000				

※「著作物と推定されるもの」とは本調査で権利の所在が推定できるもの

※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

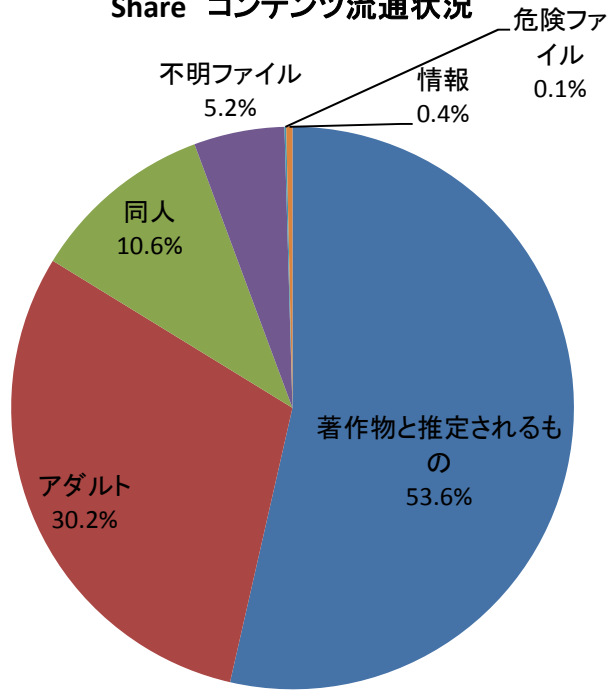
※「不明ファイル」とはタイトルからコンテンツの内容が判別できないもの

※「危険ファイル」とはタイトル、拡張子からウイルスなどと推定されるもの

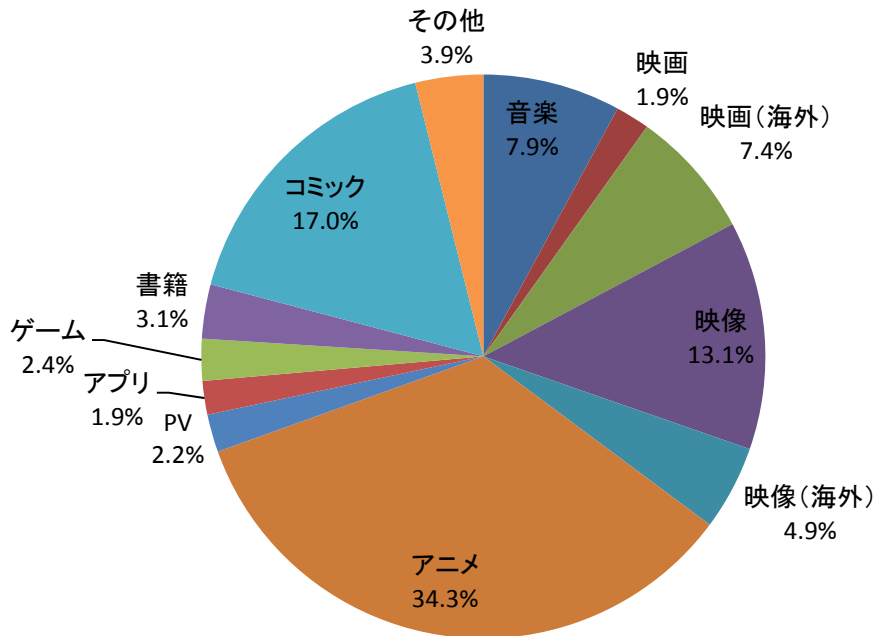
※「情報」とはウイルス感染などで流出した個人・組織等の情報だと推定されるもの

※「その他」とはコンテンツの分類が音楽、映像関連、プログラム、書籍関連に含まれないもの

Share コンテンツ流通状況



Share 著作権と推定されるコンテンツの内訳

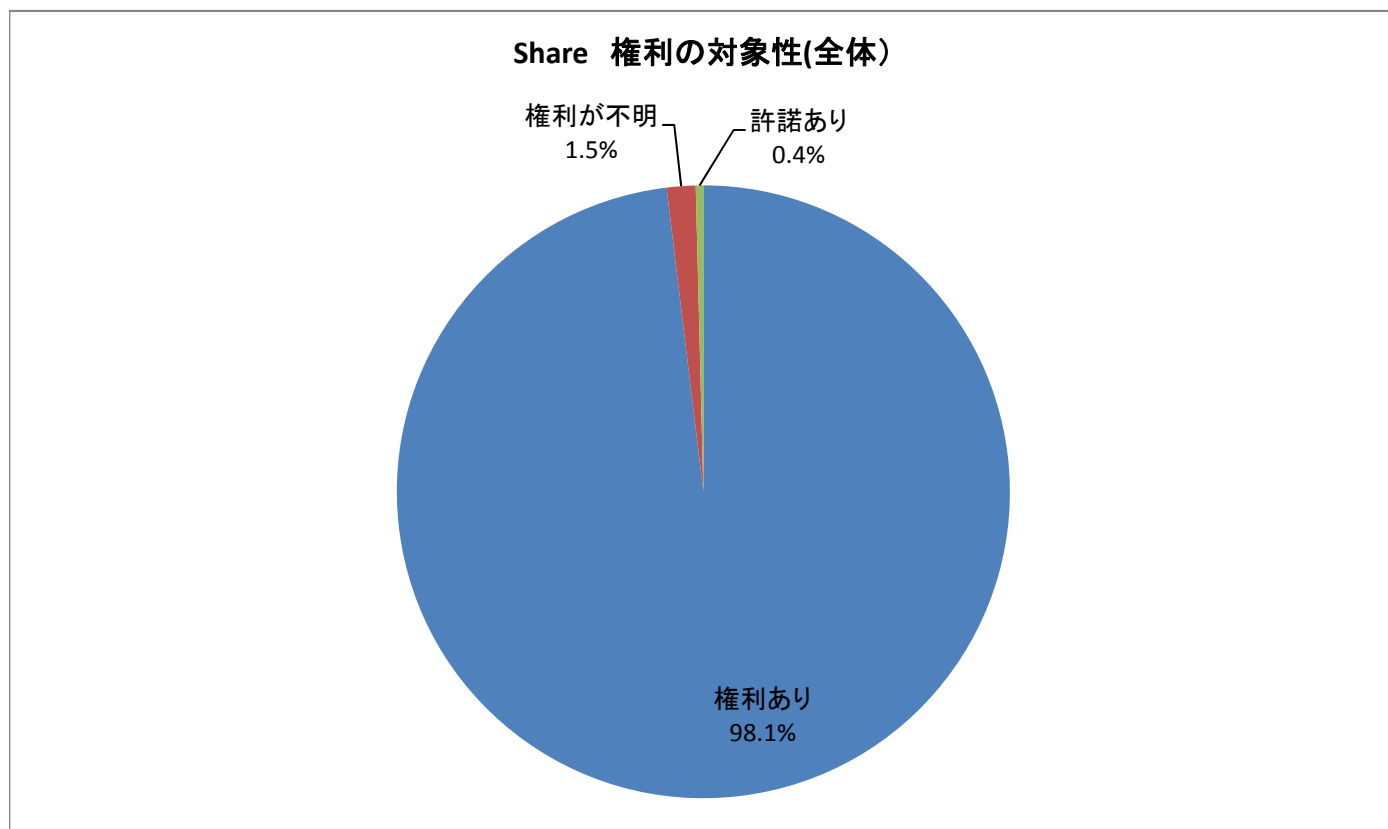


(2) 権利の対象性

① 全体

著作物と推定されるもののうち、98.1%に権利があり、かつ許諾がないものと推定される。

権利があり、許諾がないと推定されるもの	10,509
権利の所在、許諾が不明なもの	162
許諾あり、もしくは必要ないもの	43

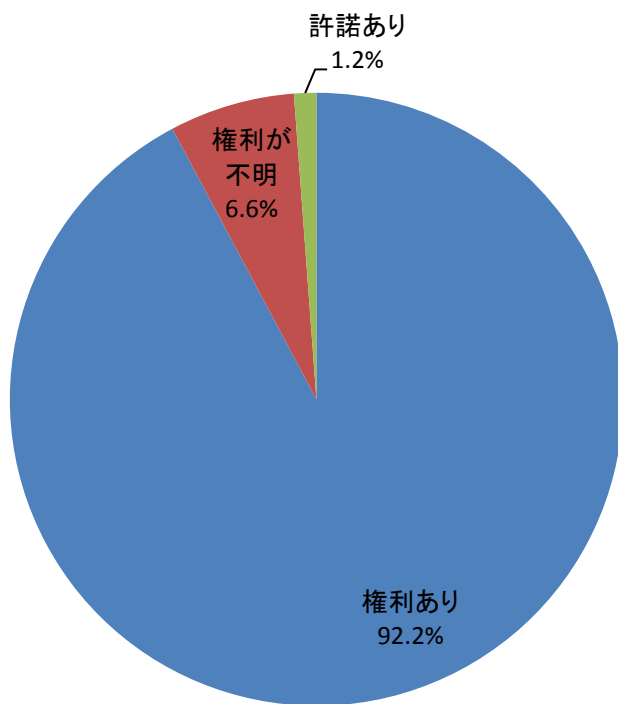


② ジャンルごと

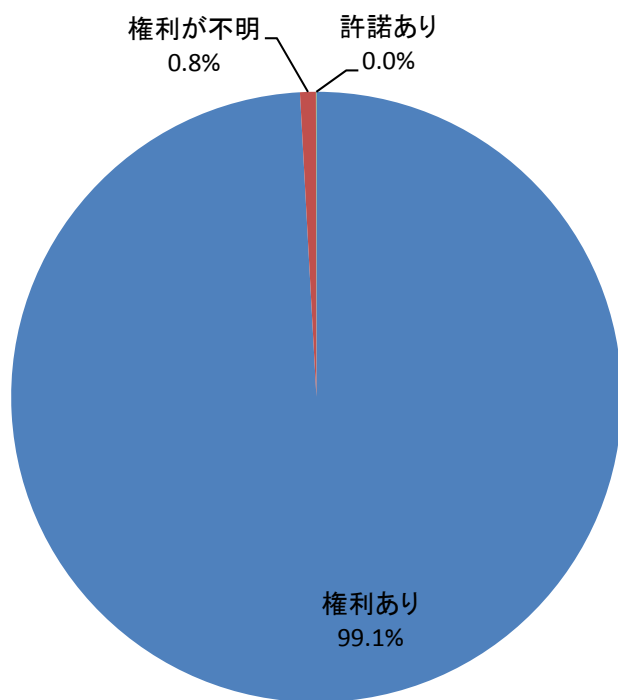
権利の対象性について、コンテンツジャンルごとに集計した。全ジャンルでほぼ90%以上が権利の対象であり、無許諾で送信されていると推定されるものであった。

ジャンル	権利あり	権利不明	許諾あり
音楽	781	56	10
映像	6,774	57	2
プログラム	415	18	30
書籍	2,155	0	0
その他	384	31	1

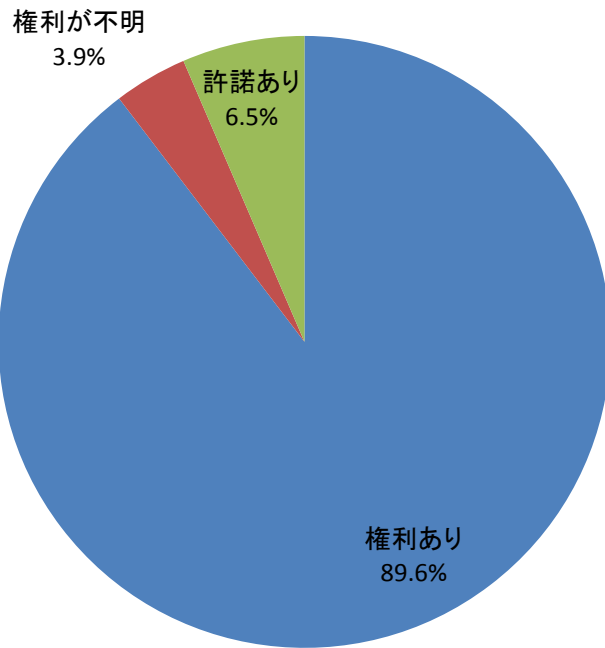
Share 権利の対象性(音楽)



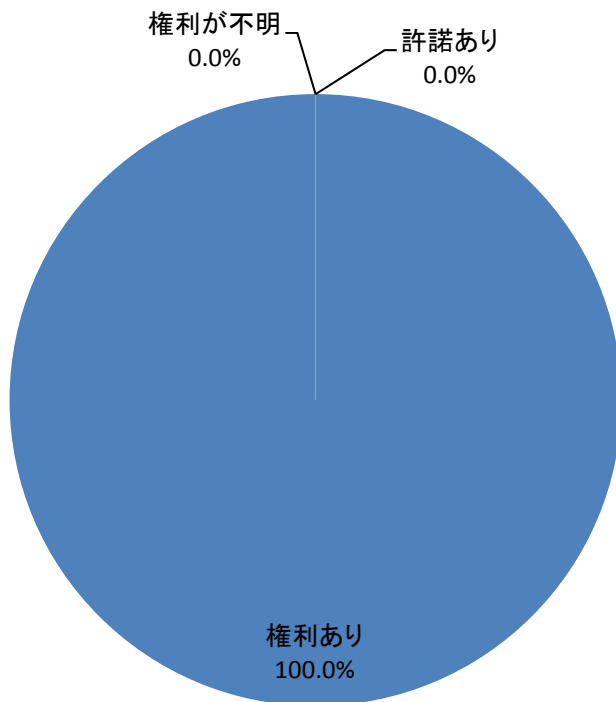
Share 権利の対象性(映像)



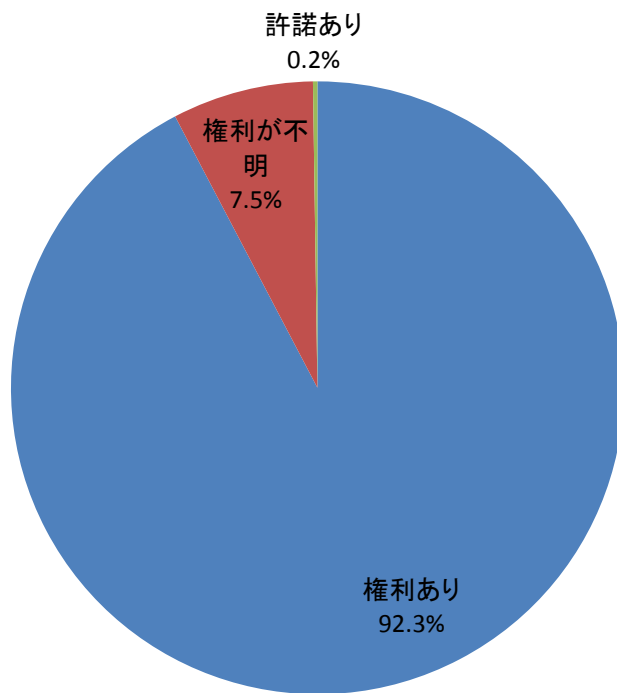
Share 権利の対象性(プログラム)



Share 権利の対象性(書籍)

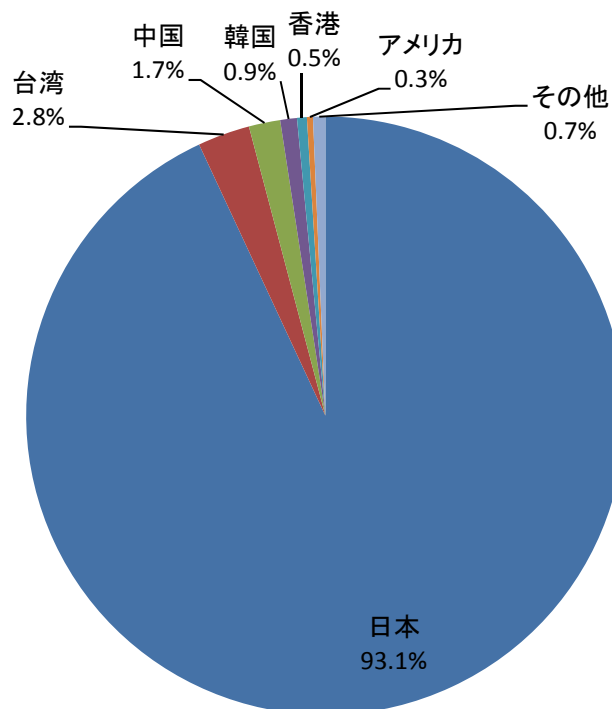


Share 権利の対象性(その他)



(3) 検出ノードの国・地域

Share 検出ノード国・地域別分布



3.PerfectDark

本調査でファイルネームと化身IDの一意的な組み合わせをノードの量として算出したところ、全数として約5.6万ノードと推定される。

(1)コンテンツの流通状況

流通コンテンツのうち、65%のコンテンツが著作物と推定される。

著作物と推定されるもの	13,008	音楽	1,737		
		映像	7,096	映画	68
				映画(海外)	123
				映像	1,601
				映像(海外)	427
				アニメ	4,674
				PV	203
		プログラム	775	アプリ	257
				ゲーム	518
		書籍	2,209	書籍	449
コミック	1,760				
その他	1,191				
アダルト	4,195				
同人	2,092				
不明ファイル	663				
危険ファイル	2				
情報	40				
計	20,000				

※「著作物と推定されるもの」とは本調査で権利の所在が推定できるもの

※「アダルト」、「同人」とは本調査で権利の所在が判別できないため、権利の対象に関する調査は見送ったもの

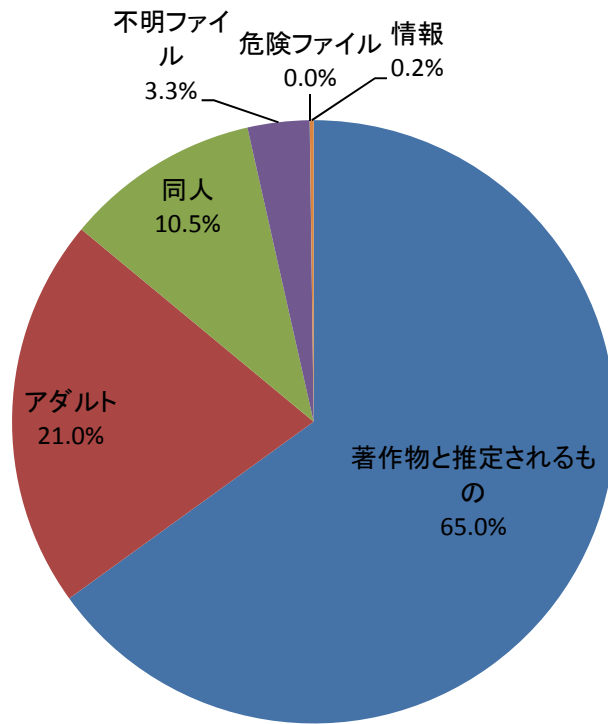
※「不明ファイル」とはタイトルからコンテンツの内容が判別できないもの

※「危険ファイル」とはタイトル、拡張子からウィルスなどと推定されるもの

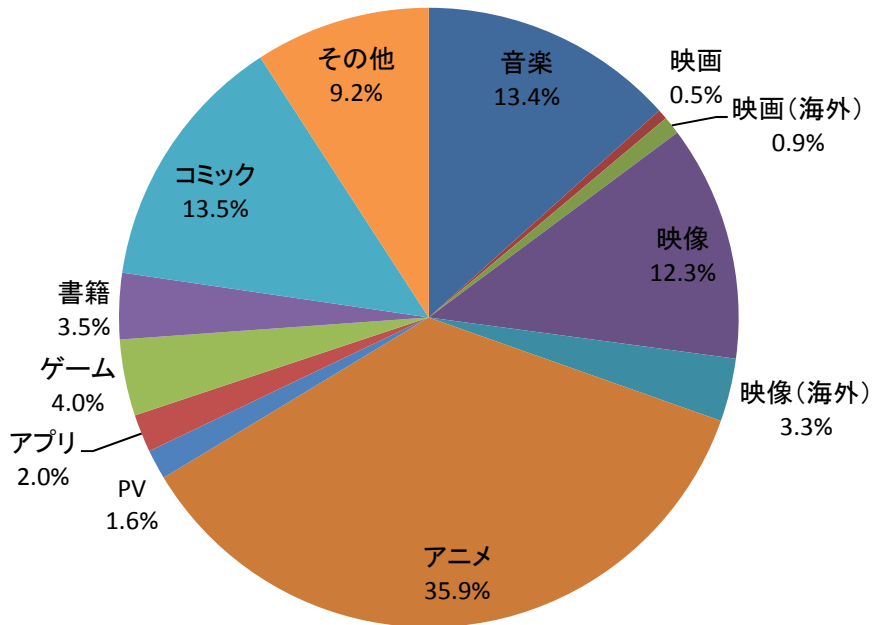
※「情報」とはウィルス感染などで流出した個人・組織等の情報だと推定されるもの

※「その他」とはコンテンツの分類が音楽、映像関連、プログラム、書籍関連に含まれないもの

PerfectDark コンテンツ流通状況



PerfectDark 著作物と推測されるコンテンツの内訳

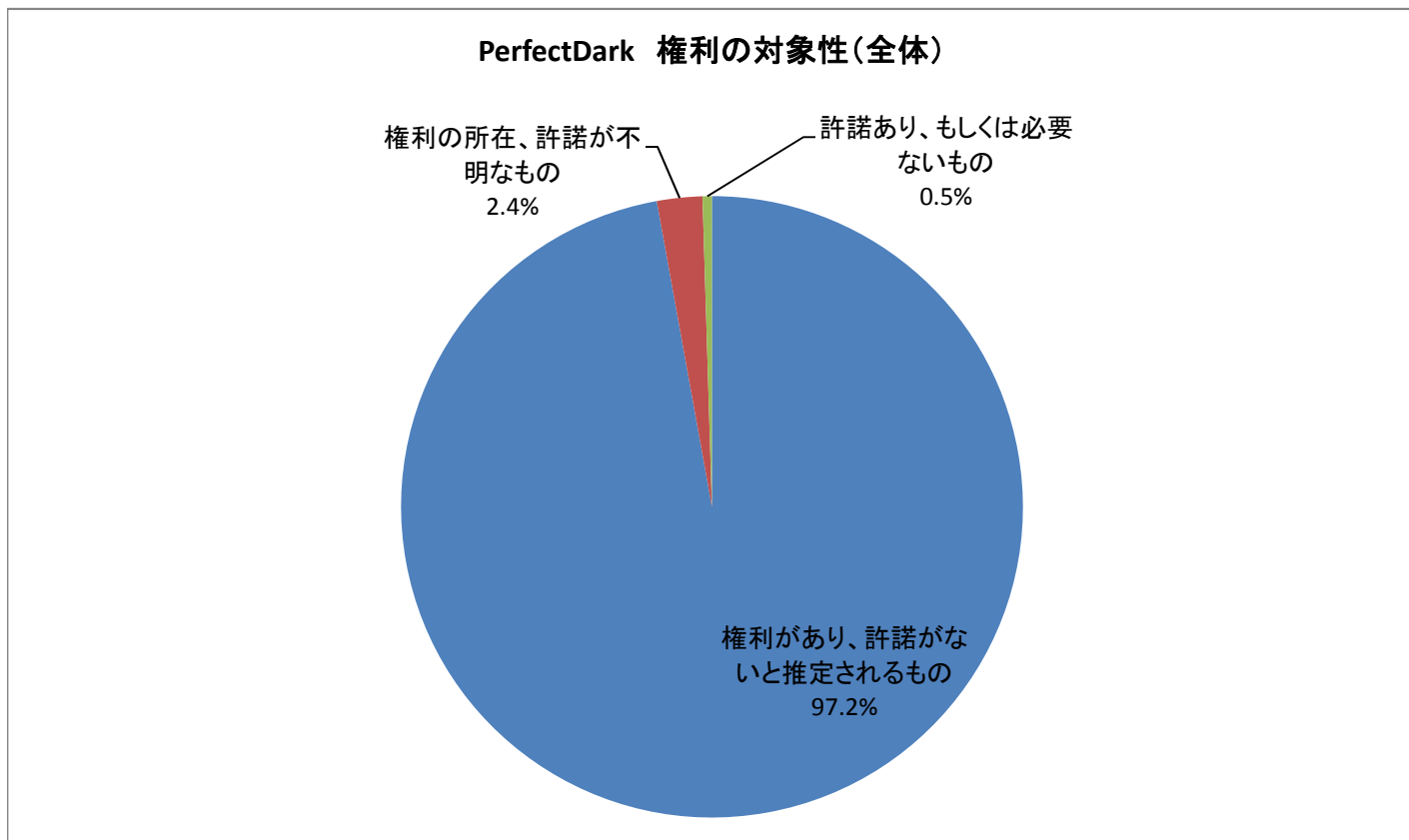


(2) 権利の対象性

① 全体

著作物と推定されるもののうち、97.2%に権利があり、かつ許諾がないものと推定される。

権利があり、許諾がないと推定されるもの	12,638
権利の所在、許諾が不明なもの	307
許諾あり、もしくは必要ないもの	63

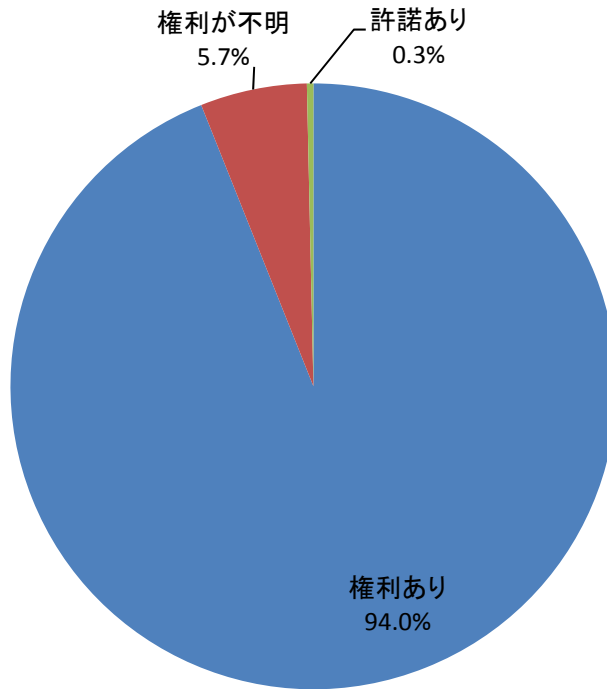


② ジャンルごと

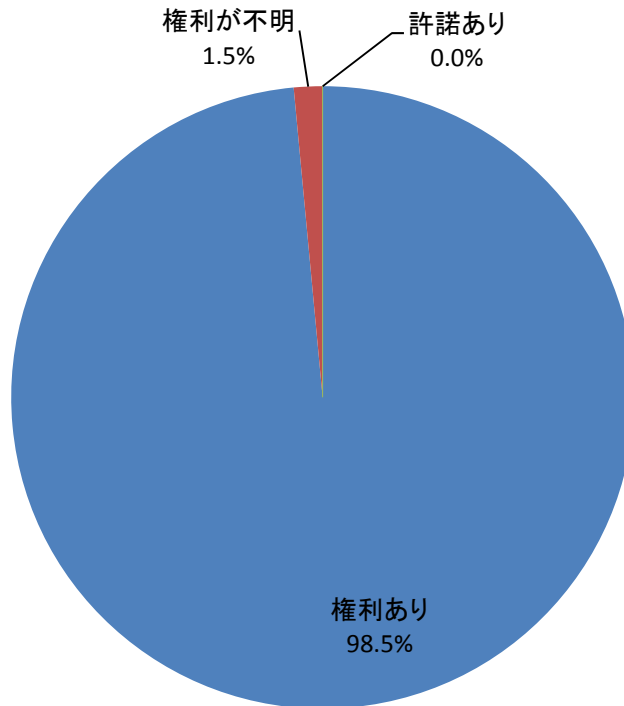
権利の対象性について、コンテンツジャンルごとに集計した。全ジャンルでほぼ90%以上が権利の対象であり、無許諾で送信されていると推定されるものであった。

ジャンル	権利あり、許諾がないと推定されるもの	権利の所在、許諾が不明なもの	許諾あり、もしくは必要ないもの
音楽	1,632	99	6
映像	6,991	105	0
プログラム	685	34	56
書籍	2,208	1	0
その他	1,122	68	1

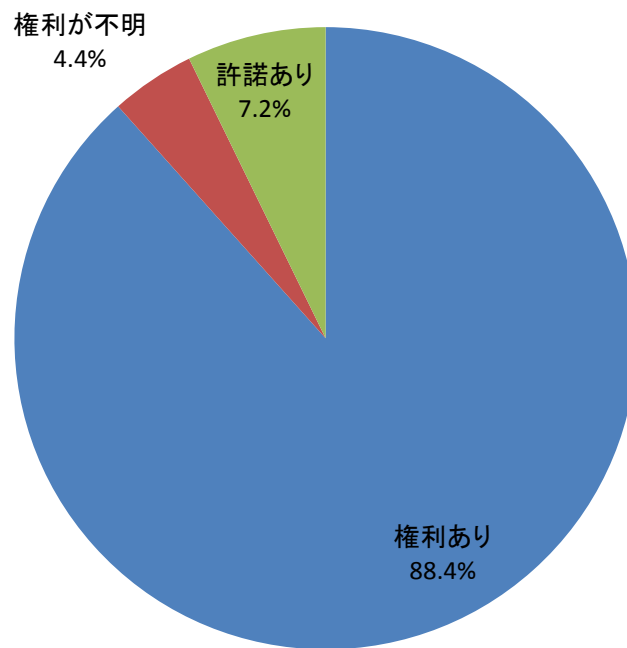
PerfectDark 権利の対象性(音楽)



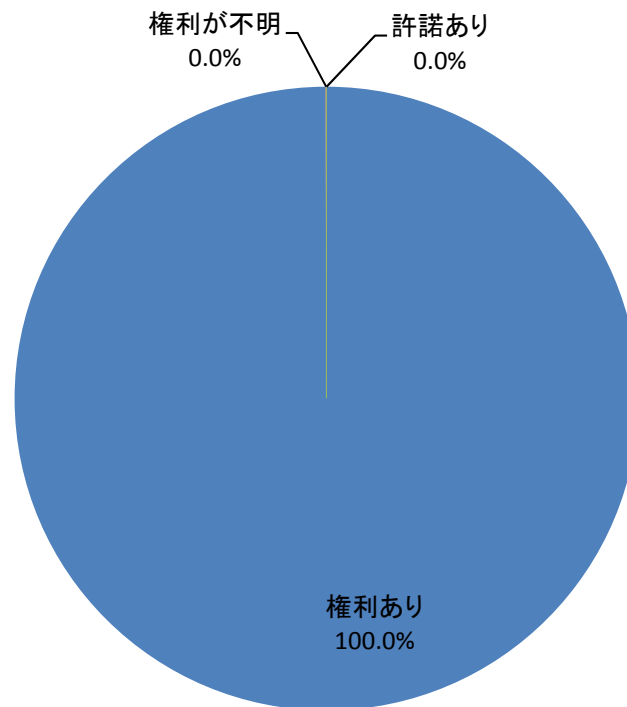
PerfectDark 権利の対象性(映像)



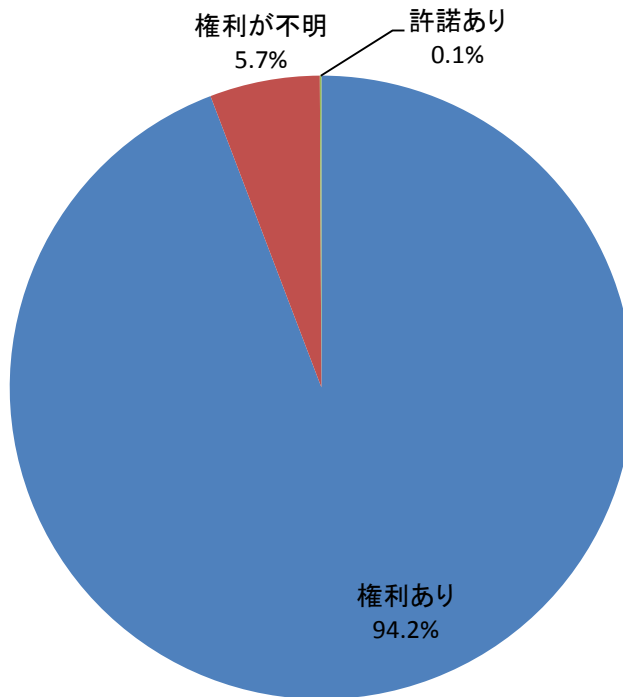
PerfectDark 権利の対象性(プログラム)



PerfectDark 権利の対象性(書籍)



PerfectDark 権利の対象性(その他)



(3) 検出ノードの国・地域

PerfectDark 検出ノード国・地域別分布

