



Introduction to **Blueprint Data**

Summary

Blueprint Data is a software company that has developed a highly detailed “human reviewed” database of contextual information that covers over 800 million URL’s / Web pages, in multiple languages, accounting for over 98% of the World Wide Web traffic (Neilson/Net Rating). To date we have used our proprietary database to design and develop specialized data security products and provide an Internet Filtering data set to companies for inclusion in integrated Internet security solutions. **Blueprint Data’s** URL Description Database can add significant value to a company security products at a cost point notably lower than our competitors.

Blueprint Data Data Store

Since 1998, **Blueprint Data** has built the world’s largest data store of Web site content and context information gathered exclusively by trained and certified Web site review specialists. **Blueprint Data’s** content specialists have evaluated hundreds of thousands of Web sites representing over 800 million Web pages, accounting for well over 98% of the World Wide Web traffic (Neilson/Net Rating) making it **the largest and most accurate data store of reviewed Web sites in the world.**

Blueprint Data includes English and Spanish Language URL reviews in its database. A small number of other sites in 10 Languages are also reviewed; however they do not have significant coverage of the language¹. Additional languages can be added as part of Service Level Agreements or as requested by customers and business partners. **Blueprint Data** uses the web surfing history and web site requests from its customer’s subscribers and users to build its URL database. Web sites that are added to the URL review queue if the URL is not a current part of the system.

The **Blueprint Data** offering provides extremely granular details on each web site, resulting in improved accuracy. Instead of the high error rates often associated with automated technologies, the approach used by most URL information companies, **Blueprint Data** is close to 100% accurate. How? Because 100% of our Web site content reviews are completed by trained and certified **Blueprint Data** employees and overseen with an outstanding Quality Assurance process.

The **Blueprint Data** system can tell you, for example, if a web site or web page contains nudity **and** the context of the nudity, artistic, educational, or medical. **Blueprint Data’s list of Web site content is sorted into 162 categories** (3 General Information, 36 Descriptive Classifications, 82 General Classification categories, and 41 Language, Sex & Nudity, and Violence ratings). **Blueprint Data’s** high evaluation standards enable the company to be the most trusted Web site content authority in the world.

¹ Chinese, Dutch, French, German, Greek, Italian, Japanese, Korean, Portuguese and Russian. No systematic or specific reviews have currently been completed by **Blueprint Data** on these languages.

[BluePrint Data](#) ensures the data stays up to date by monitoring web site changes and re-reviewing more than 33% of the data each year. [BluePrint Data](#) also continually adds new Web sites by automatically sending unknown URLs that are not in our data store to our reviewers. These review requests are placed in a queue with the more popular URLs ranking higher.

The [Blueprint Data](#) system also improves contextual relevancy. For example, if a web site or web page contains violence we can tell you if it is in an artistic, educational, medical or sports context and whether the violence is directed against humans, animals, fantasy characters or objects. This context specificity is unique. The data set is also portable and can reside on your appliance or remotely accessed allowing for simple OEM integration. To update the list you (or the end user) set a timer as to how frequently it contacts our servers for the changes since the last update. Our list of Web pages are categorized and rated according to the Internet Content Rating Association (ICRA – www.icra.org) and [Blueprint Data's](#) high evaluation standards, enabling the company to be the most trusted Web site content authority in the world.

Blueprint Data Content Information

This is a list of the content information [BluePrint Data](#) review specialists gather while at a Web site or page. We classify each Web site or page based on what category its content places it in. We then gather information as to the title, description, and keywords for the Web site, portion of the site, or page. We also review the content by examining Language, Sex & Nudity, and Violence based on Internet Content Rating Association (ICRA) standards and classify the content in categories of concern ranging from Classified / Auctions to Weapons Promotion

Using these categories and classifications provides a significant competitive advantage as:

1. The Web site content information is very detailed and accurate, as it was gathered exclusively through a human review process
2. The information on each webpage includes the context for items such as language, sex & nudity, and violence
3. The web site reviewers “see” content such as Flash animations and videos that automated classification technologies may not be able to categorize.

BluePrint Data's Three (3) General Information Categories

[BluePrint Data's](#) classification specialists and editors gather information as to:

A Short Title
A Description
Keywords that reflect the content and purpose

Describing the Site

Titles

We gather a short title for the site, section of a site, or page we are reviewing (a general summary of the content). The title is usually very similar to the URL; for example, the title for www.nintendo.com is “Nintendo”. Titles tend to be a few words in most cases.

Descriptions

We gather a longer description (a general summary of the content) for the site, section of a site or page. If a description exists in the Web page meta tags we typically use it, unless it does not accurately express describe the content. In some cases, there is a good description of the site and we make only minor adjustments to the wording. In other cases, we create a brief description of the site.

For example, a good description for www.nintendo.com might be “This is the official North American site for Nintendo, maker of the Wii, Nintendo Dual Screen, and games. Read news and reviews about games for Nintendo video game systems and see previews of games and review game codes. Home to Mario Brothers, Donkey Kong, and other video game characters.”

Descriptions of the site, section of a site, or page convey enough information so that a reader is able to understand the purpose and content of the Web site or page. In the example above, the description is 49 words long. This is slightly shorter than the average description,

Keywords

Keywords are also gathered. If keywords are embedded into the Web sites meta tags they are used unless they are deliberately misleading. From this meta tag list and from our content review specialists review and examination of the sites content, we create a list of keywords someone searching for this site would be likely type into a search engine.

For Nintendo our keywords might be: “Wii, Nintendo Dual Screen, Game Boy, Donkey Kong, Game Cube, Gaming, Mario Brothers, Pokemon, Video Games, Nintendo DS, Wii fit”.

BluePrint Data’s Thirty Six (36) Descriptive Classifications

1. Anonymous Proxies
2. Bogs / E-mail (web based) / Message Boards
3. Chat / Instant Communications (web based – typically port 80)
4. Classified / Auctions
5. Drugs / Alcohol /Tobacco Advocacy, or Promotion
6. Entertainment
7. Fraud / Cheating/Illegal Activities
8. Gambling
9. Games
10. Gay / Lesbian
11. Intimate Apparel
12. Hate Speech / Intolerance / Discrimination
13. Higher Education / Colleges / Universities
14. Investing / Stocks
15. Job Search / Careers / Continuing Education
16. Motor Vehicles
17. Mysticism / Astrology
18. News
19. Personals / Dating / Romance
20. Personal Web Content (personal web sites) / User Generated Content
(social networking sites)
21. Politics
22. Religion
23. Sex Education - Preteen
24. Sex Education - Teen / Advanced (including abortion)
25. Shopping
26. Sports / Recreation
27. Suicide Promotion

- 28. Travel / Tourism
- 29. Weapons Promotion or Sale

Special Categories Children

- 30. Child - The web site content is educational or primarily designed and targeted to children.
- 31. Bad Example - Material that might be perceived as setting a bad example for young children
- 32. Disturbing - Material that might disturb young children (cults, autopsies, etc)
- 33. Education - Content of high educational value for children

Special Categories Content

- 34. Domain for Sale
- 35. Under Construction / no content
- 36. Unreachable / Dead / no content

BluePrint Data's Eighty Two (82) General Classification Category or Categories

BluePrint Data classification specialists and editors ask themselves the question, "If I were to search for this site at a search engine, what category or categories would I search under?"

1. Action/Adventure

- a. Sports/Recreation
- b. Transportation/Vehicles
- c. Travel (destinations, vacations)
- d. Other

2. Art/Entertainment

- a. Art (painting, sculpture)
- b. Books
- c. Boy's sites (identified specifically as such in title or content)
- d. Comics
- e. Contests
- f. Dance
- g. Games/Puzzles
- h. Girl's sites (identified specifically as such in title or content)
- i. Hobbies (crafts, collecting)
- j. Literature (literary fiction, poetry)
- k. Museums
- l. Music
- m. Radio
- n. Storytelling & Puppetry
- o. Television
- p. Theatre
- q. Toys
- r. Other

3. Business

- a. Businesses (specific company sites)
- b. Insurance
- c. Investing
- d. Retirement
- e. Other

4. Culture/Society

- a. Cities/Towns
- b. Cultures of the World
- c. Fashion/Style
- d. Food and eating (preparation, eateries)
- e. Holidays
- f. News
- g. Religion/Spirituality
- h. Other

5. Lifestyle

- a. Fitness
- b. Health/Nutrition
- c. Home/Garden
- d. Kids (topics for kids)
- e. Parenting
- f. Pets
- g. Self-help/Psychology (relationships, human behavior, dating)
- h. Seniors
- i. Teenager sites (ages 13 – 19)
- j. Other

6. Education/Research

- a. Academic Studies (general subjects, curriculum)
- b. Books (academic subjects, learning)
- c. Business
- d. Careers
- e. Colleges/Universities
- f. Economics
- g. Government
- h. History
- i. Homework/Studying
- j. Kindergarten through 12th grade
- k. Language arts
- l. Law
- m. Math
- n. Politics (political organizations, campaigns)
- o. Reference (almanacs, dictionaries, encyclopedias)
- p. Teaching (help for teachers, training)
- q. Other

7. Science/Technology

- a. Animals
- b. Astronomy/Space
- c. Biology
- d. Computers (clipart, hardware, programming)
- e. Dinosaurs
- f. Earth sciences (weather, oceans)
- g. Ecology (environment, eco-systems)
- h. Geography and Maps

- i. Internet (net services, Web ware, support)
- j. Medicine
- k. Science (mathematics, physics, chemistry)
- l. Technology (engineering, building, manufacturing)
- m. Other

Additional Information

BluePrint Data's Forty One (41) ICRA Standard Classifications Reviewed for:

Language, Sex & Nudity, and Violence

Language

Explicit sexual language

Crude words or profanity

Mild expletives

None, no language (expletives, crude or profane words, or sexual language)

Context - This material:

- appears in a context intended to be artistic
- appears in a context intended to be educational
- appears in a context intended to be medical

Sexually Explicit - Mild Sexuality - No Sexual Content

Adult/Sexually Explicit (often referred to as hard core pornography)

- Erections or female genitals in detail
- Explicit sexual acts
- Obscured or implied sexual acts
- Visible sexual touching

Nudity and Sexual Material (often referred to as soft core pornography or erotica)

- Male genitals
- Female genitals
- Female breasts
- Bare buttocks
- Passionate kissing

No nudity or sexual material

Context - This material:

- appears in a context intended to be artistic
- appears in a context intended to be educational
- appears in a context intended to be medical

Violence – Directed to Humans – Directed to Animals – Directed to Fantasy Characters

Sexual violence / rape

Blood and gore, human beings

Blood and gore, animals

Blood and gore, fantasy characters (including animation)

Killing of human beings
Killing of animals
Killing of fantasy characters (including animation)
Deliberate injury to human beings
Deliberate injury to animals
Deliberate injury to fantasy characters (including animation)
Deliberate damage to objects

None, no violent material

Context - This material:

- appears in a context intended to be artistic
- appears in a context intended to be educational
- appears in a context intended to be medical
- only appears in a sports related context where the violence - depicts aggression allowed within the rules for that sport

Note: Sports related violence that appears outside the realm of the rules of a sport is not considered violence in a sports related context. Sites dedicated primarily to showing or depicting illegal fighting or violence outside the realm of the game rules or are intentional in their display and glorification of violence (actions not allowed by rules of the sport) would be rated based on the content shown and are NOT part of the sports related context classification.

We hope this document adequately expresses the extent of the information on Web site content gathered as part of the [BluePrint Data's](#) URL reviews undertaken strictly by our highly trained human reviewers. Please feel free to contact us anytime for more information or with any comments, questions, or suggestions you may have.

Contact Information

Bob Dahlstrom
CEO
Bob@blueprintdata.com
+1-904-398-1911 United States (Florida)

Kevin Bespolka
Business Development
kevin@blueprintdata.com
+44 (0) 1962 868-922 United Kingdom (London)

Lars Bespolka
Managing Director, Asia-Pacific
Lars@blueprintdata.com
+65 9855 1463 Mobile – Asia Pacific

Neeraj Kumar
Director of Sales, India
indiasales@blueprintdata.com
Office +91-1128032496
+91-9350800393 or +91-9350107003 (Mobile)
New Delhi