

Glossary of Key Terms

AAPOR: American Association for Public Opinion Research.

ABM: Agent-based Modelling, a computational method for simulating social behaviour.

Access Grid: See 'e-Social Science'.

Algorithm: A process or set of well-defined rules to be followed in calculations or other problem-solving operations.

Analytics: The systematic computational analysis of data or statistics; the results therefrom.

Answer piping: In online surveys, the facility to insert the response from a previous question into a question appearing later in the survey.

Application Programming Interface ('API'): A source code interface that a computer application, operating system or 'library' (in computer science, a collection of sub- programs used to develop software) provides to support requests for services to be made of it by a computer program. One function of APIs is to interact with databases that render HTML (see separate entry) code. Such interfaces enable other computer applications to interact with survey software (or other kinds of software).

Artificial Intelligence ('AI'): The scientific understanding of the mechanisms underlying thought and intelligent behaviour, and their embodiment in machines.

AoIR: Association of Internet Researchers.

Autoethnography: Accounts of the subjective experience of the ethnographic researcher being embedded in Internet activities in diverse contexts, directed to understanding digital sociality.

ACA: Automated Content Analysis.

Avatar: A representation of a human, animal or other animate object enabling the representation's participation in some form of online interaction.

Beeper studies: Experiential time-sampling research whereby participants report by various means their activities in progress at the time a signal is activated by a device carried on or about the person. Responses were originally entered on a paper instrument, but more recently include online response modes. Also known as the Experience Sampling Method.

Big Data: Online data marked by volume, velocity and variety.

BBM: Blackberry Messenger, a cell phone with strong encryption facilities.

Blog: A diary-like genre in which the 'blogger' records and/or comments on their own activity/beliefs and/or that of others, often including perspectives on current events, posting the 'blog' on the Web. May include audio and visual information as well as text, and the opportunity to comment on what is posted.

Blog platform: Software with which to write and post blogs; popular platforms include Blogger, WordPress, Tumblr, LiveJournal, Medium and Weebly.

Blogosphere: Blogs considered collectively along with their writers and readers as a distinct online network.

Bulletin board: An Internet site where users can post comments about a particular issue or topic and reply to other users' postings.

CAQDAS: Computer Assisted Qualitative Data Analysis; software for the analysis of qualitative data, chiefly text, but also audio, video and still images.

CartoDB: Suite of online tools that can integrate datasets from social media platforms, marketing datasets and GIS tools, and that provides features enabling users to produce choropleth and animated torque maps and to apply Cascading Style Sheets and Structured Query Language editing panels to the maps.

CASIC: Computer Assisted Survey Information Collection.

Centroid: Central point within a mapped area of interest.

Chatroom: An online communications environment facilitating discussion between subscribed members.

Chloropleth map: A map which uses differences in shade, colour or the placing of symbols within set areas to indicate the average values of a particular quantity in those areas.

Click-stream analysis: Analysis of how users negotiate a path around a website.

Click-through data: Data relating to the process of a visitor clicking on a web page and going to a web site whose link was provided on the page; in common usage it applies to web advertisements. The 'click-through rate' is the ratio of users who click on a specific link to the number of total users who view a page.

Client-side: Computer resources such as programs or information that are held on the user's computer rather than on or from the server to which the computer is linked. See also 'server-side'.

Cloud, The: Servers leased from large provider corporations.

Collaboratory, collaboratories: Distributed research groups that work together via online technologies enabling data exchange, communication and real-time collaboration over data transmitted across networks.

ColorBrewer: Online tool for selecting perceptually graded colour schemes; <http://colorbrewer.org>.

Common Gateway Interface ('CGI'): A scripting language that allows various commands to be executed by the researcher's web server based on the actions of the respondent.

Common Logfile Format ('CLF'): Hits defined as web elements transferred from a server to the user's browser. Common 'log fields' include the user's IP address, timestamp from the server, request for the element or web page, the status of the request, and the number of bytes transferred.

Community detection: The search for, and discovery of, coherent groups in a given dataset; in social network analysis a ‘community’ is a subset of nodes in a network that displays denser connectivity internally than externally.

Computational grid: See ‘e-Social Science’.

Computer-Supported Cooperative Work: A field of social and behavioural science concerned with the ways in which people apply and relate to information technologies when they are mutually engaged in tasks using those technologies.

Conversation Analysis (CA): Analytic approach allied to ethnomethodology; both are concerned with the real-time production of social order.

Cookie: A small file, automatically placed on a user’s computer, enabling the unique identification of browsers and users’ hypertext pathways.

Co-presence: Interaction between social actors that takes place in the same physical space, rather than being computer-mediated.

Cosplay: For ‘costume play’, where costumes or fashion inspired by movies or comic strips are worn during the playing of an online game.

Coverage error: When some part of a relevant population cannot be included in a survey sample.

Crowdsourcing: Databases assembled through collective effort for the public good, often by large numbers of individuals.

Cyberethnography: Ethnography conducted in online environments.

DARPA: Defense Advanced Research Projects Administration, a US federal government agency.

DDI: Data Documentation Initiative, a metadata standard used by social science data archives.

Data dump: Download of a dataset from a server.

Data grid: See 'e-Social Science'.

Data integration: A computational process enabling the linking together of different datasets.

Data mining: A set of procedures such as clustering and pattern-recognition algorithms that search large datasets for patterns. It is usually atheoretical, using unsupervised learning and identifying patterns in data and summarising them without reference to a conceptual or theoretical organising framework.

Data poisoning: Providing wrong or mis-leading information. A main type is subject fraud, especially where a survey instrument is forwarded to individuals outside the intended sample.

Data subject: Those who are subject of personal data and enjoy specified rights in respect of such data.

Data warehouse: Subject-oriented, integrated collection of data.

DOI: Digital Object Identifier; a unique numerical identifier assigned to an online entity.

Digital Rights Management: Professional and legal regulation of legitimate access to, and use of, digital resources.

Digital trace: Indicator of human activity created in the course of online interaction, e.g., patterns of search behaviour apparent from web log files (see 'web log').

DIVER: Software package that enables manipulation of panoramic views on scenes in a digital video.

Documentality: Extent to which data used in a study are recorded and available post hoc, ideally including description of research design, how data collection proceeded in practice and characteristics of the data.

Drop-out: Withdrawal of research subjects from participating in a research study, especially web surveys.

Edge: An undirected relationship between nodes in a network. See also 'tie'.

Emoji: Graphical icon used to convey expressive feeling.

Emoticon: A figurative representation designed to display a writer's mood or emotion and formed using only characters available on a QWERTY keyboard; most common is the 'smiley' [:-)], first used in 1982.

Encryption: Procedures for coding data in transit so that only those authorised with rights to see and use the data may do so.

End User License: Conditions and rights associated with the use of online datasets and other resources.

Entertainment poll: Surveys conducted for their amusement value. On the Internet they largely consist of websites where any visitor can respond to a posted survey. As unscientific as are telephone call-in polls.

e-Social Science: A range of computational resources and procedures using Grid and High Performance Computing to facilitate social science research, comprising Access Grid (support for using online video conferencing), Computational Grid (support for computation of very large and/or complex requirements) and Data Grid (support for discovery, collation and transfer of distributed datasets). In natural science, the equivalent term is 'e-Science'. The term 'e-Research' is also in use as a generic alternative to subject- specific terminology. In the US, the term in use is 'cyber-research'. See also 'Grid'.

Event history model: A model for analysing the occurrence of a succession of similar events, like marriages or jobs.

Experience Sampling Method: see Beeper Studies.

Expert systems: A sub-field of 'artificial intelligence' (see separate entry) that attempts to enable computers to perform a task as well as human experts by using an 'ontology' (see separate entry) for a substantive domain to reason about it.

Extensible Mark-up Language ('XML'): A flexible text format that is used for data exchange. This general purpose markup language is designed to be readable by humans, while also providing metadata tags for

content that can be easily recognised by computers.

fff: For face-to-face; also shown as ‘f2f’.

File Transfer Protocol: A protocol enabling computers and servers to transmit data across networks.

Firewall: A means of providing security of Internet user accounts. There are both hardware and software firewalls. None is 100% effective against hackers (those seeking illegitimate access to users’ accounts).

Folksonomy: Also called ‘collaborative tagging’ or ‘social tagging’. The practice and method of collaboratively creating and managing tags to annotate and categorise content. Usually, freely chosen keywords are used rather than a controlled vocabulary.

Fuzzy Logic: Approach to reasoning based on degrees of truth rather than the binary logic of true/false.

Gamification: The application of the elements of game playing to other areas of activity.

Geographical Information System (or Geographic Information System): Software handling geographical information and its visual representation. See also ‘raster data’ and ‘vector data’. ‘Geographic Information Science’ is the sub-discipline of geography concerned with the application of computational techniques.

Geotagging: Assigning a geographical/spatial location to online entities such as social media posts.

Geoweb: Geo-spatial affordances of Web 2.0.

GPS: Global Positioning System (also known as ‘Navstar’); a satellite based system for finding locations, a fundamental component of the ‘geo-referencing’ of a digital entity.

Google Hangouts for Work: An online video and text communication application.

Granularity: The fineness or coarseness of the detail available from a given data source.

Graph API: An application that accesses the Facebook social graph.

Grid: A distributed computing infrastructure that combines parallel and distributed computer platforms to enable computational operations exceeding the capacities of individual desktop computers.

Griefing: Irritating other players during online game play.

Grounded Theory: Qualitative analysis method involving close engagement with data, in-vivo coding and careful development of codes and categories that are authentic to research participants rather than being imposed by the researchers' preconceptions.

GUI: Graphical User Interface, the user's 'window' on a modern software program, comprising the program's screen display, operational iconography and tools, and using symbols rather than text to request actions and execute commands.

Harvested e-mail: Sets of e-mail addresses collected from postings on the web and from individuals knowingly or not knowingly solicited for their e-mail address.

Heuristics: Rule of thumb that works most of the time but does not provide a provably true solution.

HD Video: High Definition Video.

Homophily: In network analysis, the degree to which individuals of like type are prone to link to each other.

Human-Computer Interaction ('HCI'): A field of social and behavioural science concerned with the ways that people apply and relate to computer technologies.

Human Subjects Model (also called 'Human Subjects Research Model'): A model of ethical guidelines developed in reaction against scientific practice in Nazi Germany. Its key elements are the protection of confidentiality, anonymity and the use of informed consent.

Hyperlink: A user-assigned or automatically generated connection between two or more points on online

documents or other online artefacts.

Hypertext: An unstructured series of pages and links between pages in a network.

Hypertext Mark Up Language (HTML): A standard for marking up documents containing text and multimedia objects and linking those documents. Initial basis of the World Wide Web.

Hypertext Transfer Protocol (HTTP): A text-based protocol that is commonly used for transferring information across the Internet.

Institutional Review Board: A body charged with determining that the potential risks to research subjects are outweighed by the potential benefits of the research. Also called 'ethics committees'.

Intelligent agent: A software program possessing a form of 'artificial intelligence' (see separate entry) sufficient to sense changes in a complex environment and act on them to achieve goals on behalf of users.

Intercept survey: Pop-up surveys that use systematic sampling for every k th visitor to a website or page.

IBNIS: Internet-based Neighbourhood Information Systems; geo-spatial tools that classify populations into discrete geodemographic clusters.

ICT: Information and Communication Technologies.

ID chipping: Sensors medically inserted subcutaneously to enable personal identity to be tracked.

Indexicality: The idea that the meaning to participants of any action is inseparable from the immediate context of its production.

IK (Indigenous Knowledge): Knowledge from cultures outside the Western tradition of scholarship.

Interoperability: Procedures and computer programs enabling the linking of datasets to facilitate analytic inquiries that cannot be satisfied by reference to a single dataset. Involves the assignment of 'metadata'

tags to archived data.

IMR: Internet-mediated Research.

IOT or IoT (Internet of Things): Internet-enabled non-IRCT devices such as consumer appliances.

IP Address (Internet Protocol Address): The ‘address’ is the identifying number of the computer from which a given Internet transaction has taken place. Internet Service Providers may assign addresses ‘dynamically’ so that two sessions from the same machine show different numbers.

IRCT: Internet and Related Communications Technologies.

Internet Relay Chat: An instant messaging protocol for online communication.

IUP: Internet User Population.

Interoperable, interoperability: The ability to exchange and use information between devices or systems.

JSON object: An object in a Java Script Object Notation, a common data interchange format.

Java: A platform-independent programming language. A Java ‘applet’ can be included in a web page; the applet’s code is transferred to the user’s browser, which then executes the code. Java is suited to use in complex survey instruments. Since some users disable Java, researchers may also use HTML in tandem with CGI to present interactive forms on the web. See also ‘HTML’, ‘Common Gateway Interface’.

KML: Keyhole Markup Language, an XML notation for expressing geographic annotation and visualization.

KWIC: Key Word in Context, a fundamental element of content analysis methodology.

KBS/Knowledge Based System: A computer system based on domain knowledge.

KDD: Knowledge Discovery from Data.

Latent Dirichlet Allocation: A hierarchical Bayesian technique automating discovery of topics from texts.

Logfile analyser: Programs that use log file data to generate summary statistics on the number of page requests, the domains, the time spent on a website or on particular documents, and so on. Prominent log file analyzers include AWStats, Analog, and Deep Log Analyzer.

Mapbox: Open Source online mapping tool; others include QGIS and Leaflet.

Mash-up: A collation and correlation of information from a variety of online sources, often quickly done to form a first overview of information available on a topic. Also the name for a combination of short pieces of video or audio from multiple sources into a single new product.

MOOC: Massive Open Online Course.

MMOG: Massively Multi-player Online Game.

Measurement error: When survey response differs from the ‘true’ response, for example, because respondents have not candidly answered sensitive questions; see ‘Social desirability effect’.

Mechanical Turk: Online site run by Amazon offering those wishing to conduct surveys a pool of paid volunteers from which to sample, the per-participant cost being less than standard incentive payments.

Metadata: Data about data. May include references to schemas, provenance and information quality.

Middleware: Software components that are deployed together with existing software systems on the user’s computer platform in order to provide generic services between those systems. The principal use is in data integration, where the software tools are designed to reconcile descriptive and format differences between datasets and/or other computational entities to allow their unimpeded interaction (‘interoperability’). See also ‘interoperability’, ‘wrapper’.

Multi-modal: A research design combining several different modes of data collection.

Multiplex (as in 'multiplex network'): A network that includes edges of multiple types.

Multi-sited: Ethnographic fieldwork involving several different sites connected only analytically.

NSF: The US National Science Foundation.

Natural Language Processing ('NLP'): A subfield of 'artificial intelligence' (see separate entry) in which computer software is used to automatically generate and understand natural human language. Natural language *generation* systems convert information from databases into normal-sounding language, while natural language *understanding* systems convert normal language into more formal representations of knowledge that a computer can manipulate.

NeoGeography: Embedding of geographical information through online technologies and practices.

Netnography: Modified ethnographic techniques developed to enable efficient study of online domains, often deployed in a marketing context and conducted solely online.

Netiquette: Norms of appropriate online behaviour.

Neural network: In computing, an algorithm ('see separate entry') that attempts to mimic human reasoning by linking a series of artificial neurons to one another that are exposed to inputs and generate outputs, with a view to creating an adaptive system capable of learning to solve problems.

Newsgroup: An online forum enabling discussion between subscribers.

Node: Individual actors, people or things in a network.

NodeXL: User-friendly social media data collector that automatically downloads data in a format suitable for network analysis and provides a variety of tools for network analysis.

Non-reactive data: Data collected for research purposes without the subject of the data being aware that it is being collected. Also called 'unobtrusive data'.

Ontology: In computer science, a knowledge base that holds semantic relationships between terms and is used to reason about a substantive domain. In computer science, ontologies generally consist of a 'semantic network' linking individual objects, classes of objects, attributes or features describing those objects, and relationships between objects. This meaning of the term is distinct from its usage in philosophy.

Open data/Open government: Initiatives that make digital data collected by government freely available online to download.

OGC (Open Geospatial Consortium): A standards-setting organization that seeks to develop the potential of geospatial content for commercial and government use.

Open Source: Software whose source code is made freely available by the programmer so that others may customise it and/or elaborate its functionality.

OCR: Optical Character Recognition.

Opt-in panel: Individuals who have volunteered to participate in an ongoing web-based survey or series of web-based surveys, often following a solicitation on a website.

Opt-out: A source of bias that occurs when survey sample members choose not to participate in a survey.

Opt-in samples also cause bias due to lack of information about those who chose not to opt in.

OWL: Web Ontology Language often used on the World Wide Web.

Panel survey: Survey drawing on a panel of participants pre-recruited via a probability sampling method; different sub-sets of participants are sampled according to specific survey topic.

Paradata: Data about the process of data collection, in an online survey context including information like the amount of time to answer a particular question.

PAR: Participatory Action Research, a research technique where subjects actively participate in the

research as equals with the researcher; usually one goal is to empower the subjects through the research.

Petabyte: 1,000 terabytes.

Phishing: Masquerading as a trustworthy entity in an electronic communication in order to fraudulently acquire sensitive personal information, such as passwords and bank card details.

Podcast: A digital media file, or a series of such files, that is distributed over the Internet using syndication feeds for playback on portable media players and personal computers. The term can refer either to the content itself or to the method by which it is syndicated.

Pop up: An associated link that appears in a small window on the users screen when visiting a website, often used to invite users to respond to an online survey.

PDF: Portable Document Format.

Portal: An Internet site providing access or links to other sites.

Prosumption: The idea that consumers increasingly also produce the things that they in turn consume.

Purposive sampling: Sampling using non-probability methods.

Python: The de facto standard programming language.

QGIS: A free GIS software program.

QR: Quick Response codes, which can be scanned with a QR reader to receive product information or event information additional to the information displayed on an object showing a QR code.

Radio button: Response display for web survey participants; the participant clicks on a graphical representation of a 'button' applying to their preferred response (these resemble those on 1950s car radios).

Random Digit Dialing ('RDD'): A random sampling method used in telephone surveys.

Raster data: Pixel-based geographical data. Also see ‘Geographical Information System’ and ‘Vector data’.

Really Simple Syndication: See ‘RSS’.

RTSP: Real Time Transfer Protocol, a network protocol to control streaming media servers.

Relational database: A database that maintains a set of separate, related files (tables), but combines data elements from the files for queries and reports when required. Such databases are organised around a data table in which a row refers to a single case and a column refers to a specific attribute.

Reproducibility Project: Study that seeks to reproduce the findings of published studies across the span of natural and social science disciplines.

Resource Description Format (‘RDF’): A format providing the means of representing relationships between elements of document content.

Resource discovery: Location of datasets satisfying an analytic requirement from repositories and archives, particularly in an online environment.

Revenge porn: The release to a wider audience of pornographic images originally exchanged by parties to an intimate relationship and intended to harm the subject of the image after the relationship has ceased.

RSS: Rich Site Summary (also often referred to as Really Simple Syndication): format for displaying a summary or complete text of frequently updated content from a web site.

Scalable, scalability/scaleability: The ability of a network, process or system to process an increasing amount of work or to be enlarged such that it can handle greater loads.

Scanner (strictly, ‘image scanner’): A device that scans optical images, printed text, handwriting and objects and stores the results in a file that they can be used online; 3-D scanners simulate this function in three dimensions.

Scraper: Automated computer scripts that parse web page content so it is useful as data. See also ‘spider’.

Script: Pieces of computer code that can be assembled with each other to solve computational problems.

Search Engine: Software with which computer users can search for required information on online networks, notably the Internet and World Wide Web. Major providers like Google, Yahoo! and MSN collect data on searches for unspecified uses that are retained for unspecified periods of time.

Secondary analysis: Analysis of data conducted by those other than the original collectors of the data.

Second Life: An online virtual world which had one million regular users by 2013.

Secure Socket Layer: A communication protocol whose primary goal is to provide private and reliable communication between two computer applications.

Seed set: A set of web pages purposively selected to satisfy a query. See also ‘scraper’ and ‘spider’.

Semantic Map: Maps of words visually displaying meaning-based connections between words/phrases.

Semantic Web: The use of formal knowledge representation techniques to comment and annotate web resources. Its key feature is the imposition of structure on data to facilitate automated processing. Unlike the Internet, in the Semantic Web, resources are commented and annotated explicitly to form expressive knowledge bases, enabling automation of various procedures in application to data resources. Autonomous computer programs (‘agents’) can access the content for selective retrieval and analysis.

Sensor: A miniaturised device that records and transmits an animate being’s or object’s proximity, motion, skin temperature, brain and nerve activity, and anything else that can be measured.

Sensor studies: Research involving the collection of data from sensors applied to people or objects (‘tagging’) to track behavioural information, including transactions. Also called ‘remote sensor studies’.

Sentiment analysis: A suite of methods for understanding the current attitudes and intentions of a subject

population, including both lexical and machine learning approaches, whose main tasks include polarity detection, sentiment strength detection, and fine-grained emotion detection.

Server Log File: An automated system recording network search terms and date and time of requests.

Server-side: Computer resources such as programs or information that is not held on the user's computer but on or from the server to which the computer is linked. See also 'client-side'.

Skip logic: In online surveys the ability to go to a question based on the response given to a previous question.

Social affordance: A feature of a system that enables a form of social interaction.

Social cartography: Cartographic information practices based on non-expert epistemologies and lay users of interactive mapping technologies, platforms, and software.

Social desirability effect: When research subjects provide responses that they think the researcher wants to hear or that they think put them in a good light, rather than their actual views or behaviour.

Social media: A generic reference to social network sites like Facebook or LinkedIn.

Social Network Analysis: Techniques for the analysis of social networks and their role in social behaviour. Not confined to, but greatly enhanced by, online information and resources.

SNS: Social Network Site like Facebook or LinkedIn (see Social Media).

Social shaping: A social science concept referring to the role that technology has in shaping society and social relations and the way that society and social relations affect the development and application of technologies.

Social Simulation: A social science method using computer models that recreate the dynamics of human actions and interactions.

Spam: Unwanted online communication, usually received via e-mail and generally containing advertising. Named after a 1970s' Monty Python TV sketch in which a restaurant offered only Spam, a formed meat product sold in cans.

Spider: A special class of scraper that follows links between web pages and collects information along the way. Data for spiders often comes from a seed set. See also 'scraper' and 'seed set'.

SPSS: Statistical Package for the Social Sciences, a statistical software product.

Structured Query Language ('SQL'): A standard language for inserting data into a database, selecting subsets, aggregating results and producing 'reports'.

Supervised machine learning: The machine learning task of inferring a function from labeled training data.

Tablet: A handheld, touchscreen computer.

Terabyte: One million million bytes.

Text mining: Techniques employing computer applications to analyse large volumes of text in order to identify patterns, concordances (associations) and links between hitherto unrelated information sources in the public domain.

3-D Printer: A device for making a physical object from a three-dimensional digital model.

Tie: Relationship between nodes in a network. See 'edge'.

Triangulation: The use of different research methods in combination, originally to test for 'convergent validation', but now more often to enable a fuller, richer account of the phenomenon under study. A frequent combination is of one or more quantitative and one or more qualitative methods. Also called 'mixed methods', 'multiple methods' or 'multi-method'. More broadly, the comparison of information from two or more sources to check, verify and/or validate information from a primary source.

Transactional data: Data created largely for administrative purposes and relating to online transactions, such as purchases from online retailers, or company payroll records.

URL (Uniform Resource Locator): The unique address of an Internet resource.

Usability: Level of difficulty of using an online application. Affects response rate and data quality.

Usenet: A distributed conversation system much used as source of aggregated message records for social network analysis.

Ushahidi: A crisis mapping organisation that uses the Geoweb and crowdsourcing to create activist mapping aimed at facilitating international social justice work.

Vector data: Coordinate-based geographical data. Also see ‘Geographical Information System’ and ‘Raster data’.

VRE: Virtual Reality Environment.

Visualization: Visual representation of data, often in charts and graphs, and designed to show complex statistical, numerical data in accessible visual ways; data visualization software includes Big Picture, Chart Builder, D3, Graphviz and, for specific chart types, e-Sankey.

Voice Over Internet Protocol (VOIP): Internet telephony (e.g., Skype or Video MSN), later versions of which enable users to see as well as hear each other.

VGI: Volunteered Geographic Information, such as that gathered through crowdsourcing information about a specific location.

Webcam: A video camera connected to a computer.

Web crawler: Also called a ‘Web spider’. A program or automated script which browses the World Wide Web in a methodical, automated manner. Many sites, in particular search engines, use crawlers to provide

up-to-date data. Web crawlers are mostly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches. Crawlers can also automate website maintenance tasks, e.g., checking links or validating HTML code, and gather specific types of information from web pages, e.g., harvesting e-mail addresses. See also 'Spider', 'Scraper'.

Web log: Information available from Common Logfile Format data and clickstream analysis. See entries for 'Common Logfile Format' and 'Clickstream analysis'.

Webometrics: Measures of online activity, both automated and human, that provide information about online behaviour.

Web Services: A software system designed to support interoperable machine or application-oriented interaction over a network.

WebSM: An organisation and website offering information on software for use in social science research.

Web survey: A social survey conducted over the web. Also called an 'Internet Survey'.

Web 2.0: A development of the original World Wide Web providing features promoting user participation and support for large-scale social networking applications.

WhatsApp: A free Instant Messaging service.

Wiki: A computer application enabling incremental contributions by multiple users to elaborate a resource ranging from single documents to encyclopaedias and dictionaries.

Wrapper: Components of middleware systems with which datasets can be web- enabled or Grid-enabled. They resolve dataset heterogeneity and thus enable integration of different datasets.

XML: See 'Extensible Mark Up Language'.