Website
Radar

Website-Radar 2023: Zustand österreichischer Webseiten

15. April 2023

Website-Radar

1

Einleitung

Wer sind wir, was machen wir und warum machen wir es?

2

Anforderungen an eine Website

Was, ausser Design und SEO sollte noch berücksichtigt werden?

3

Die Realisierung

Welche Technologien haben wir benutzt?

4

Die Ergebnisse

Was haben wir herausgefunden?

Website Radar

Einleitung

Wer sind wir, was machen wir und warum machen wir es?

Website-Radar / die Geschichte / die Motivation

- Ich betreue ich „nebenbei“ mehrere Websites.

- Eine davon hatte DEN Anwaltsbrief bekommen

- Die Abwicklung verschlang Ressourcen
- Die ständige Überprüfung mehrerer Websites auch

- Immer wieder „kleine Projekte“ mit den Kindern
- Die Kinder sind größer geworden, die Projekte sind größer geworden.

Eine Idee ist geboren

- Aufgrund unserer Kenntnisse im Bereich **Datenschutz**, **Computer-Sicherheit, Software Entwicklung** und **Marketing** haben wir einen eigenen Scanner programmiert, der vor allem die Aspekte Google Fonts und Cookies überprüft.

- Denn Datenschutz ist wichtig!

Website-Radar

"To run an efficient team, you only need three people: a Hipster, a Hacker, and a Hustler.“

Rei Inamoto, März 2012

Anforderungen an eine Website

Was, **außer Design und SEO** sollte noch beim Erstellen und Betreiben einer Website berücksichtigt werden?

Motivation für Sicherheit und Hosting

- (einige) Angriffsvektoren
  - Automatisierte Angriffe (e.g. Brute-Force und bekannte Schwachstellen)
  - Phishing und Social Engineering
  - Distributed Denial of Service (DDoS)
- (einige) mögliche Folgen
  - Produktivitätsausfall
  - Reputationsverlust
  - Datenverlust
- Dem steht gegenüber
  - Kosten und Aufwand

Motivation für Datenschutz

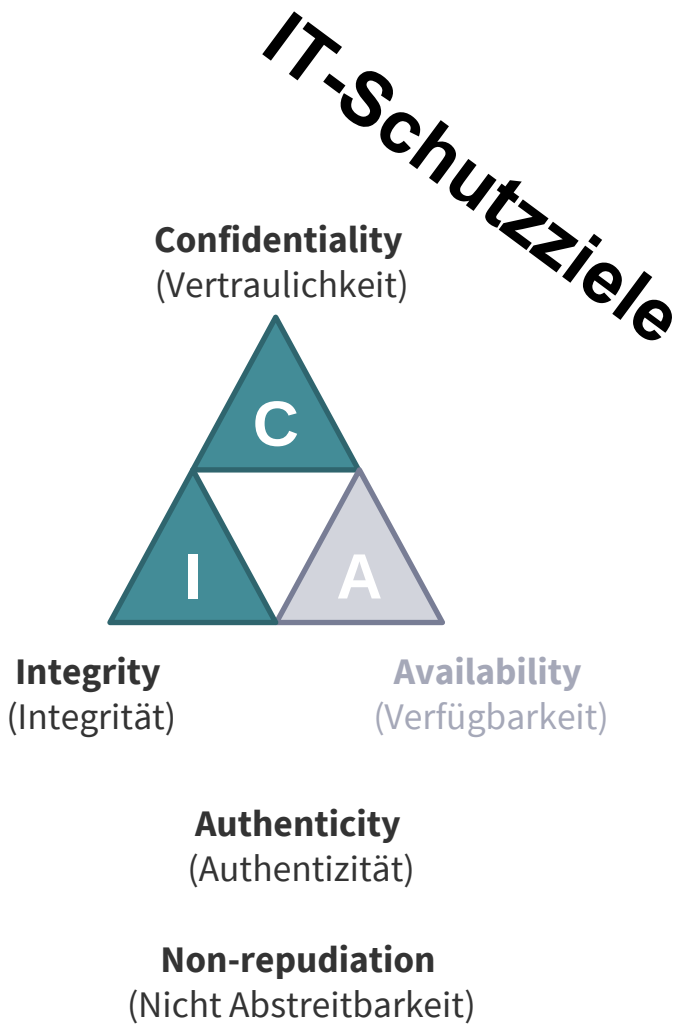
- Pull Faktoren
  - Schutz der Privatsphäre von Nutzern
  - Stärkung des Vertrauens ins eigene Unternehmen
  - Wettbewerbsvorteil
  - Einhaltung rechtlicher Verpflichtungen
- Push Faktoren
  - rechtlichen Konsequenzen
  - hohe Geldstrafen (20 Mio Euro oder 4% des weltweiten Jahresumsatzes)
  - Reputationsverlust
  - Datenverlust

Datenschutz und Rechtliches

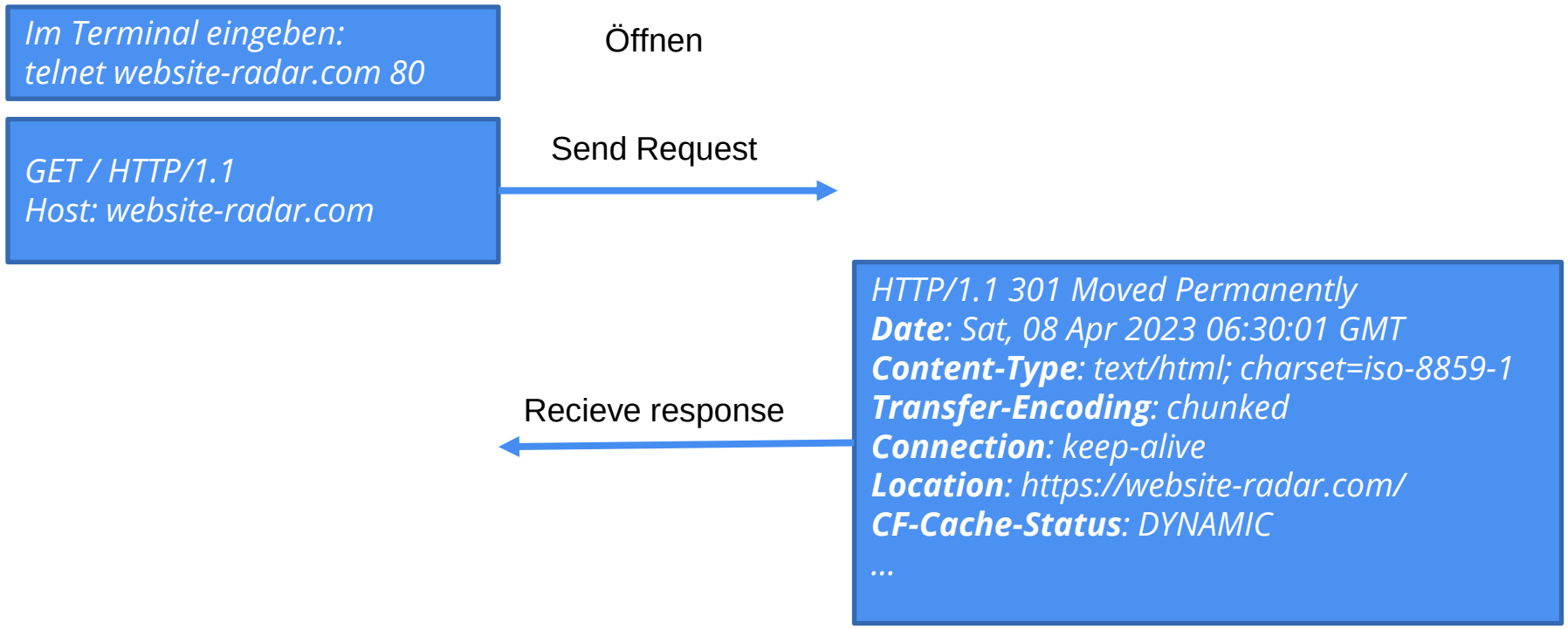
- Schutz personenbezogener Daten: der heilige Gral!
- Cookies: Einholung von Einwilligungen
- Download von digitalen Assets: Google Fonts & Co
- Eingaben: Contact-Form, Newsletter Anmeldung
- Verantwortung: Impressum
- Datenschutzerklärung
- AGB/T&C

Zertifikate

- Das Zertifikat Deiner Website ist ein „elektronischer Echtheitsausweis“
- Sorgt für einige IT-Schutzziele
  - Integrität (Daten bleiben während der Verarbeitung unversehrt, vollständig und aktuell)
  - Vertraulichkeit (Nur der Webserver sieht Deine Eingaben in der Contact Form)
  - Authentizität (als Ergänzung zur Integrität, Von wem kommen die Daten)



HTTP Security Header



Dein Server gibt dem Webbrowser Anweisungen!

HTTP Security Header

Entscheidend für Sicherheit und SEO Deiner Website

CONTENT SECURITY POLICY

Von welchen Quellen darf der Browser Inhalte laden

HTTP STRICT TRANSPORT SECURITY

Erzwingt die Verwendung von HTTPS

REFERRER-POLICY

Verhindert, dass private Infos an Dritte weitergegeben werden

PERMISSIONS-POLICY

Welche Funktionen (z.B. Standort, Kamera, Mikrofon) darf der Browser verwenden.

X-FRAME-OPTIONS

Verhindert, dass Deine Website auf anderen Webseiten verwendet wird. (Clickjacking!)

X-CONTENT-TYPE-OPTIONS

Verhindert das Manipulieren von Inhaltstypen

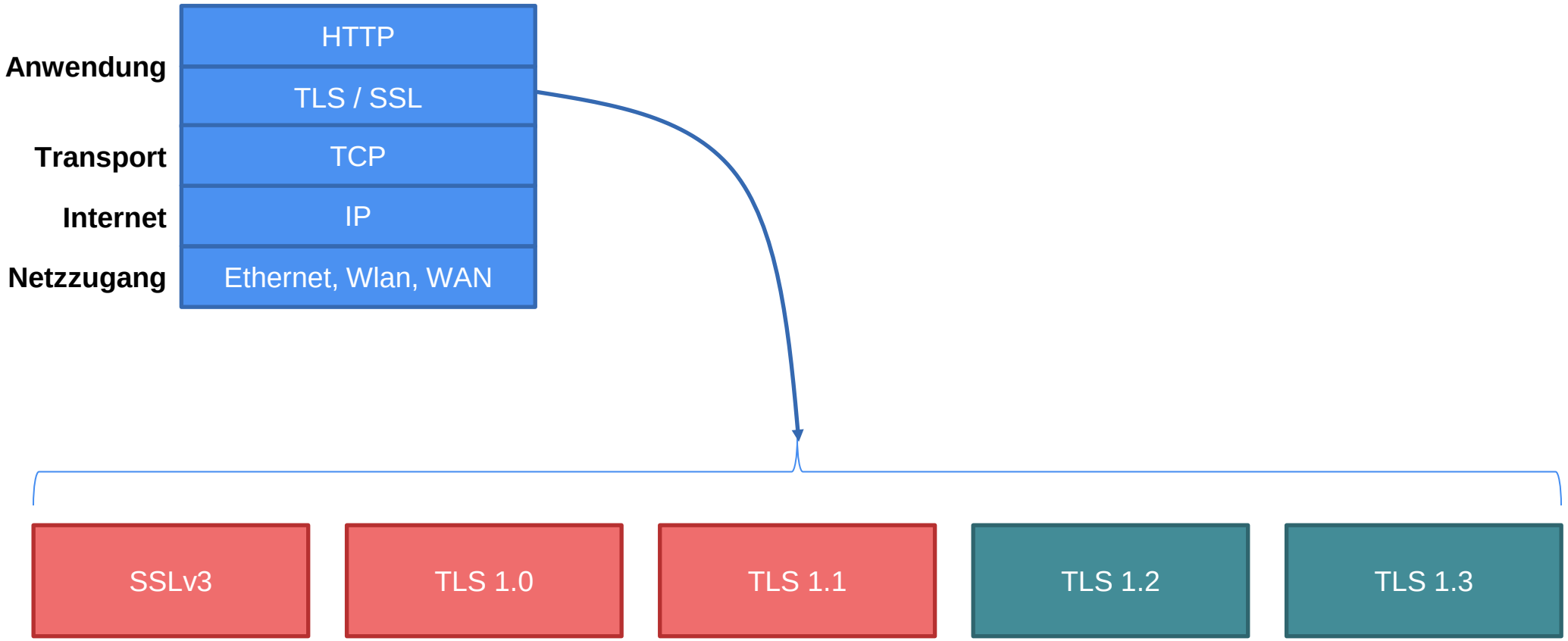
Bitte kümmert euch auch um

OSHP

https://owasp.org/www-project-secure-headers/

- O … OWASP (Open Worldwide Application Security Project)
- SHP … Secure Headers Project

Security Protokolle und Cipher Suites



Cipher Suites:
e.g. TLS_ECDH_RSA_AES256-GCM_SHA384
TLS_Schlüsselaustausch_Authentifizierung_Massenverschlüsselung_MAC

Sichere vs. nicht sichere Cipher Suites

Sicherheitsvorkehrungen

- Hosting bzw. Proxy gegen DDOS & maybe WAF
- Passwörter & MFA/2FA

- Scotty, back me up
- Patch me if you can

- Mitarbeiter Awareness & Training (Passwörter, Phishing, Datenschutz)
- Unterstützende Dokumentation (e.g. Verarbeitungsverzeichnis für Auftragsverarbeiter)
- Prozesse (e.g. Im Falle des Falles)
- Trockentraining (Backups ausprobieren)

Der Hosting Aspekt

- Webserver-Konfiguration
- Verwendete Themes
- Verwendete Plugins
- Sicherheitsvorkehrungen
- Ständige Updates (System, PHP, WordPress, Plugins)
- Regelmäßig überprüfen

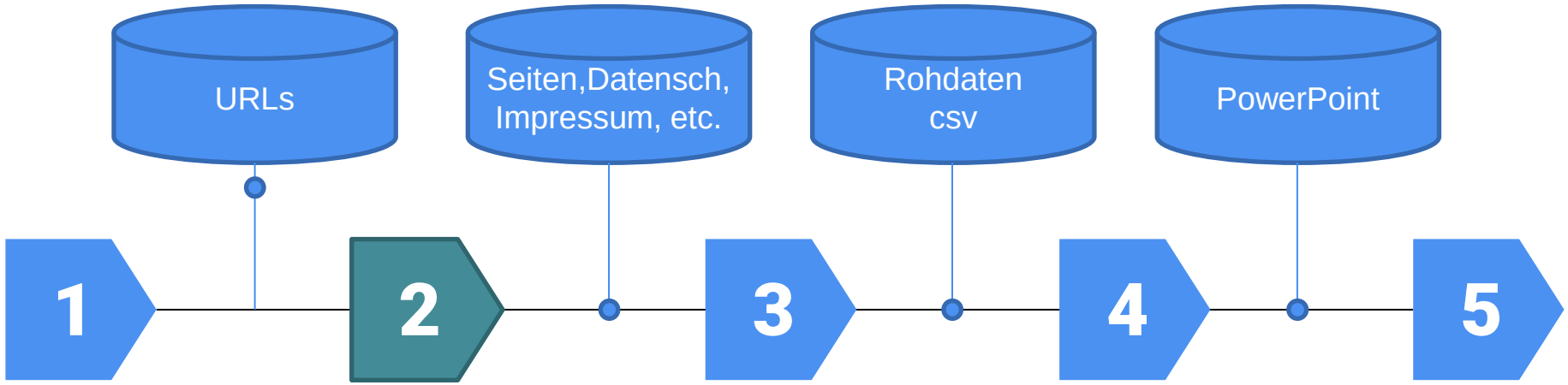
Die Realisierung

Welche Technologien haben wir benutzt?

Technology Stack

- Sprachen
  - JAVA
  - PHP: WordPress Plugins sowie Laravel Application
  - Shell Scripts, Python Scripts, REST API, JSON
- Architektur
  - Microservices Architektur
  - RabbitMQ Messaging Broker
- Weiteres
  - Um alte Sicherheitsprotokolle zu erkennen mussten Konfigurationen auf den Scanner-Hosts vorgenommen werden.

Fünf Schritte Prozess



Crawling

Sammeln und säubern von URLs

Scraping

Scannen der Seiten-Elemente

Analyzing

Analysieren der Daten

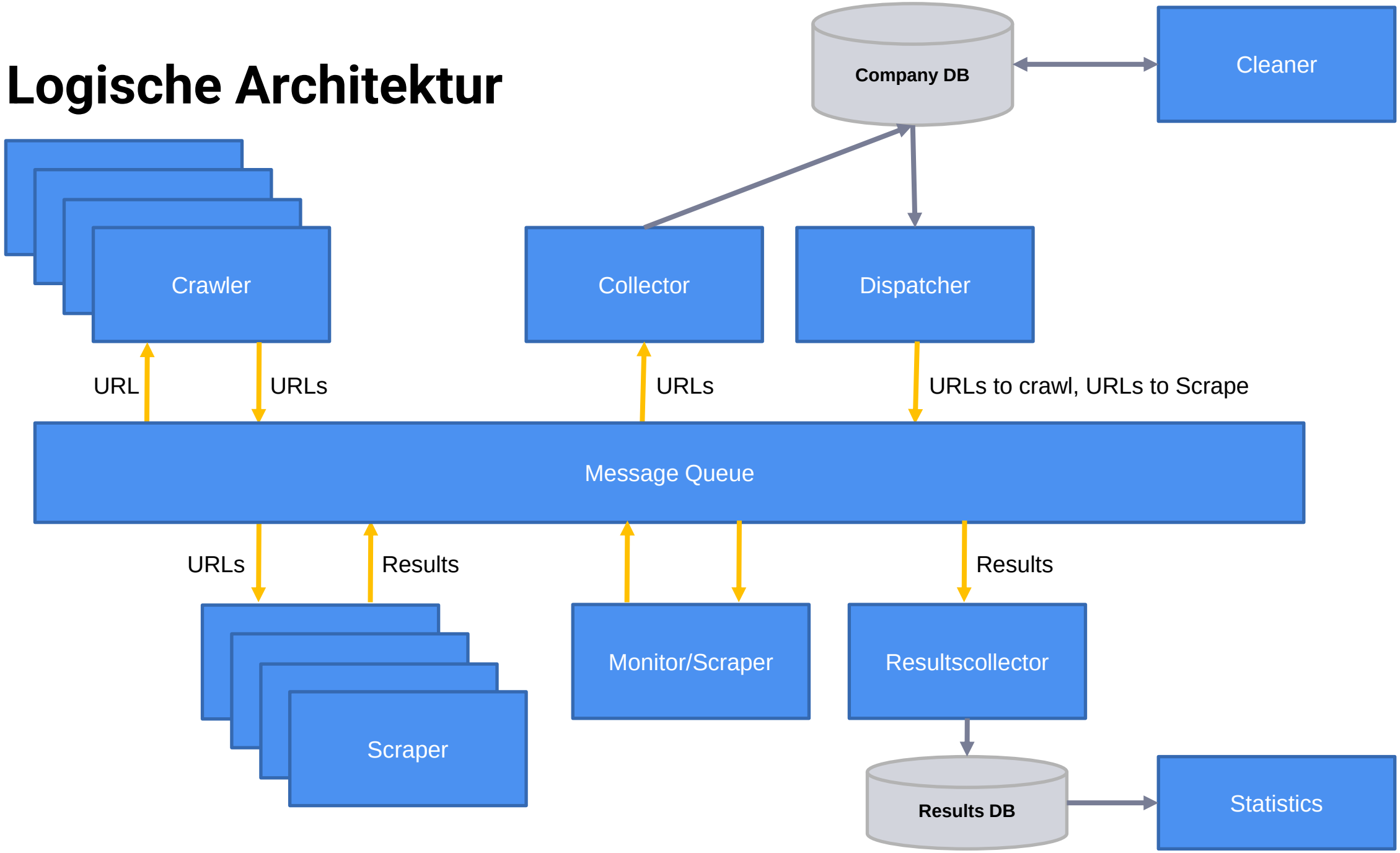
Evaluate

Auswerten und präsentieren

Improving

Unser KVP

Logische Architektur



Weshalb WordPress

- WordPress ist mit Abstand das dominante CMS in Österreich
- Als kleines Team wollten wir fokussieren und uns das größte Kuchenstück einverleiben
- Readme.txt & Lizenz Dateien, Auflistung von Verzeichnissen
- Themes, Plugins
- Versionen und CVEs abgleichen

Glossar & Blog

- In den vielen Diskussionen mit Website Besitzern haben wir festgestellt, dass in dem Kontext auch ein Glossar fehlt.
- Wir arbeiten uns langsam aber sicher in Richtung Datenschutz (auch für Firmen) und Sicherheitsaspekte (auch Cyber Security)

Website Radar

Die Ergebnisse

Was haben wir herausgefunden?

Grundlegende Daten in Österreich (nic.at)

Aktuelle Domain Zahlen in Österreich

Domain-Endung	Anzahl	davon IDN
.at	1.448.355	34.970
.co.at	33.762	497
.or.at	7.847	134
Gesamt	**1.489.964**	**35.601**

Domain-Webnutzung

Art der Webnutzung	Anzahl	Prozent
Gesperrte Domains	21.904	1,47 %
Inhalt nicht abrufbar	316.419	21,24 %
Low Content Domains	316.635	21,25 %
High Content Domains	835.006	56,04 %
Gesamt	**1.489.964**	**100,00 %**

at-Domain-Zahlen nach Inhaber-Ländern

Land	Anzahl	Prozent
Österreich	1.066.080	71,55 %
Deutschland	214.331	14,38 %
Schweiz	33.580	2,25 %
Rest der EU	92.524	6,21 %
Rest der Welt	83.449	5,60 %
Gesamt	**1.489.964**	**100,00 %**

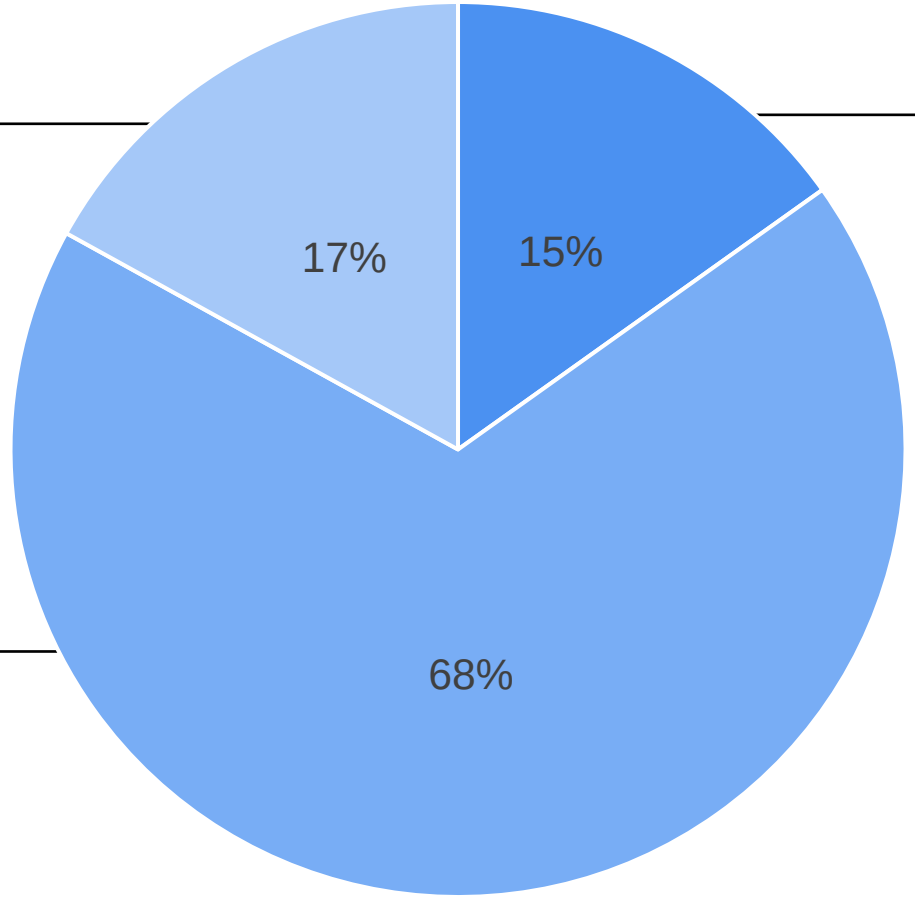
Sicherheit 1 „Nutzung von Zertifikaten“

(HTTPS)

Fehlerhafte HTTPS Verbindungen

HTTPS

Korrekte HTTPS Verbindungen



HTTP

Noch keine Zertifikate installiert.

Sicherheit 2 – „Security Protokolle“

	Relativ zu allen HTTPS Seiten
TLS 1.3	67,5%
TLS 1.2	78,7%
TLS 1.1	21,9%
TLS 1.0	20,6%
SSLv3	8,2%

Sicherheit 3 „HTTP Security Header“

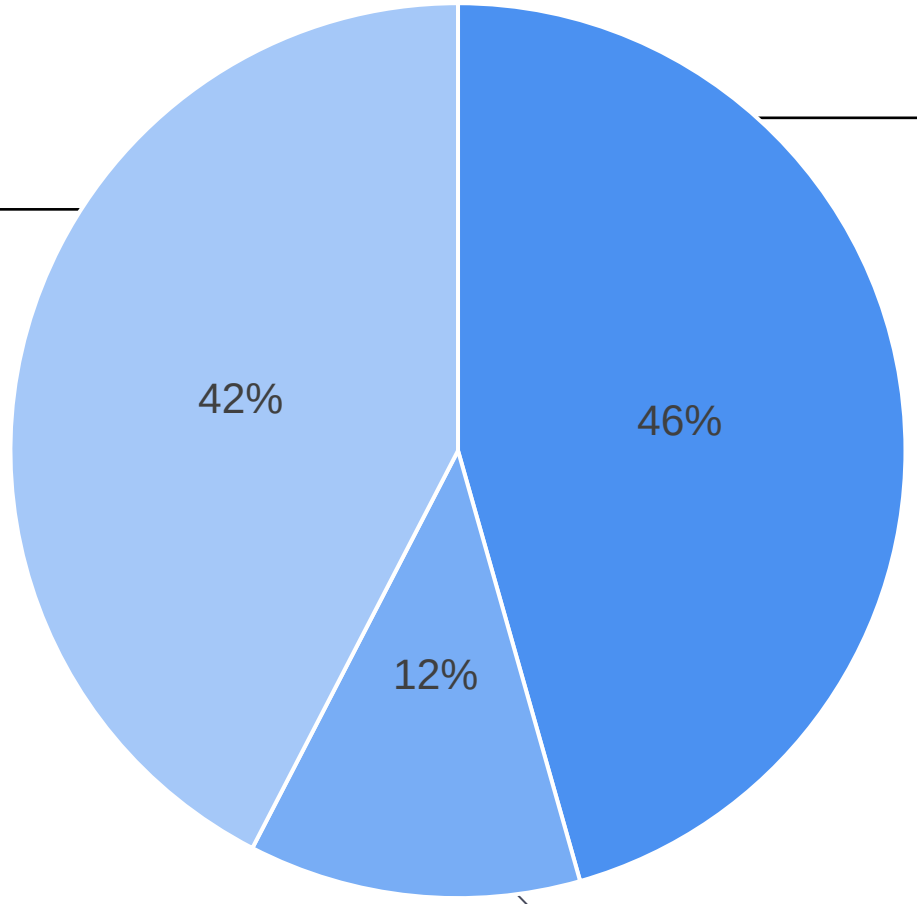
	Relativ zu allen Seiten
Content-Security-Policy	4,0%
Strict-Transport-Security	18,4%
Referrer-Policy	5,2%
Permissions-Policy	1,7%
X-Frame-Options	9,8%
X-XSS-Protection	7,4%
X-Content-Type-Options	19,5%

Cookies 1

Nicht DSGVO konform

Manche tun sich schwer!

- Ignorieren
- Falsch konfigurierte Cookie Banner
- Webshops



Die Webseite überträgt keine nicht bestätigten Cookies.

Geht doch!

Nur 1st Party Session Cookies (vor der Bestätigung)

Die beliebtesten Cookies

	Relativ zu allen Seiten
_ga	15,49%
_gid	14.09%
PHPSESSID	10.03%
_gat	5.49%
XSRF-TOKEN	4.28%
hs	4.07%
svSession	4.07%
ssr-caching	3.36%
_fbp	3.34%
fedops.logger.defaultOverrides	3.15%
__utmb	2.99%
__utmc	2.99%
__utma	2.99%
__utmz	2.99%

Google Analytics

to store and count pageviews.

to store and count pageviews.

to read and filter requests from bots.

Für XSRF – auch WIX (unsicher)

WIX Cookie aus Sicherheitsgründen

WIX

WIX Cookie für Login

WIX Cookie für Rendering

Facebook

WIX Statistics

to store time of visit

to store time of visit

store the calc. of days and time to purchase.

to store used keyword and search engine.

https://cookiedatabase.org/

Externe Digital Assets ohne Bestätigung

URL	Anteil
fonts.gstatic.com	23.45%
fonts.googleapis.com	21.3%
www.google-analytics.com	17.26%
www.googletagmanager.com	15.15%
www.google.com	14.08%
www.gstatic.com	8.09%
maps.googleapis.com	7.48%
connect.facebook.net	6.29%
www.facebook.com	5.94%
stats.g.doubleclick.net	5.84%
cdnjs.cloudflare.com	5.38%
www.youtube.com	5.18%
ajax.googleapis.com	5.14%
maps.gstatic.com	4.53%
static.wixstatic.com	4.44%
frog.wix.com	4.37%
www.google.at	4.35%
fonts.jimstatic.com	4.29%
googleads.g.doubleclick.net	4.19%
static.parastorage.com	4.19%
siteassets.parastorage.com	4.12%

Typische Fehler I

- Impressum (mit Vorsicht!)
  - Bei 19,0% aller Websites haben wir kein Impressum gefunden
- Datenschutzerklärung
  - Bei 30,7% aller Websites haben wir keine Datenschutzseite gefunden
- robots.txt
  - Bei 26,2% aller Websites haben wir keine robots.txt Datei gefunden
  - 4,8% hatten eine inkorrekte robots.txt Datei (meist 301 oder 404)
- Sitemap
  - 21,7% aller Websites haben keine (korrekte) Sitemap

Typische Fehler II

- Canonical Tag
  - Nur 60 % aller Sites haben ein Canonical Tag
  - Von denen sind 28% falsch gesetzt
- Directory Browsing
  - Bei 9,4% aller Websites konnten wir ungehindert Directories browsen
- .htaccess Datei
  - Bei 5,4% aller Websites waren die Rechte der .htaccess Datei falsch gesetzt
- WordPress readme.txt Datei
  - Bei 18,4% aller WordPress Seiten war die readme.txt Datei vorhanden
- 'Eine weitere WordPress Seite‘
  - Bei 0,6% aller WordPress Seiten war der Text noch vorhanden

Aufschlüsselung der CMS Systeme

CMS	Anteil
WordPress	60.27%
TYPO3	7.43%
Wix	6.82%
Joomla	6.33%
Jimdo	5.49%
Herold	2.38%
Contao	2.06%
Drupal	1.02%
IONOS	0.98%
Shopify	0.92%

Achtung: Bei viele Websites lässt sich das CMS (noch) nicht erkennen.

Und zuletzt …

- Umlaute in der URL (IDNs)
  - 0,15% der URLs hatten Umlaute (NIC hat ca. 2,4% registriert)
- Fonts
  - Z.B. Von allen Sites, bei denen die Fonts ohne Cookie von extern nachgeladen werden, verwenden 26% Google Fonts
- Weitere Site-Technologien wie CDN, RSS, APIs, Komprimierung, Zeichenkodierung
- Sprache(n)
- Netzwerk - IPv4 / IPv6-Adresse, Registrar, Hosting-Land
- Wer hostet die Website?

Und wer weiß …

… vielleicht werden wir das Website-Radar auch auf kommerzielle Beine stellen.

Lead Generation	Website Monitoring	Website Audits (Consulting)
Finde neue Kunden und Interessenten in unserer ständig wachsenden Datenbank. Greife schnell auf die richtigen Informationen zu. Wir fassen alles, was Du wissen musst, in einer konsolidierten Liste potenzieller Kunden zusammen: Technologie (CMS, Captcha, Cookie Dialog), Datenschutz-Status, Impressum, Hosting- und Sicherheitsaspekte, Schlüsselwörter und mehr. Mit der erweiterten Filterung kannst Du relevante Ergebnisse weiter aufschlüsseln.	Regelmäßige, **automatische** Überwachung Deiner Website bezüglich Änderungen im Datenschutz, Security, Hosting und SEO Regelmäßiges Reporting Warnungen bei problematischen Veränderungen	Untersuchen von Websites auf Einhaltung von Best-Practice für Datenschutz, Security, Hosting und SEO Identifizierung von Optimierungspotenzialen und Quick-Wins mit konkreten Handlungsempfehlungen Sowie Unterstützung bei der Umsetzung