

# IT-Infrastrukturarbeiten im Interesse der HAFL

Damit die IT-Services professionell und nachhaltig (d.h. mit minimaler Aufwand für den Betrieb und Aktualisierungen/Weiterentwicklungen) IT-Ressourcen für Lehre, Forschung und Entwicklung an der HAFL zur Verfügung stellen können, werden einige grundlegende Basisdienste benötigt. Es geht also um **Infrastruktur**-Dienste und -Dienstleistungen.

Leider ist [Infrastruktur nicht sexy](#): Wenn sie funktioniert merkt es niemand weil es als selbstverständlich vorausgesetzt wird. In diesem Dokument wollen wir aufzeigen, welche Arbeiten für eine nachhaltige und zukunftssichere Infrastruktur notwendig sind. Erst damit können zeitgemässe und notwendige Dienste im Interesse der HAFL bereitgestellt werden.

Grundsätzlich sollen den Mitarbeitenden die Nutzung von moderne IT-Werkzeugen (Hardware, Software und Services) ermöglicht werden, damit sie ihre Arbeit schneller, besser und professioneller machen können. Damit solche Werkzeuge und Services an der BFH genutzt werden können, ist einige Basisarbeit erforderlich:

- Unsere «Supercomputer» für wissenschaftliches Rechnen benötigten Software-Anpassungen, damit die vorhandenen Grafikkarten für Berechnungen benutzt werden können und um einige Grössenordnungen schneller werden. [Simon hat den Proof of Concept dafür im Sommer 2017 gemacht](#), es fehlte bisher die Zeit auf IT-Services Seite dies abzuschliessen und zu integrieren.
- Eingesetzte Programme müssen für Linux paketierrt, konfiguriert und aktuell gehalten werden. Die Konfiguration der Dienste müssen automatisiert(er) werden.
- Der Zugriff auf den grossen Datenspeicher («Ceph») läuft nur über einen Server. Fällt dieser aus, kann niemand mehr auf die Daten zugreifen und laufende Berechnungen werden abgebrochen (es dauert zwar nur ein paar Minuten bis jemand wieder automatisiert einen neuen Server erstellt, damit ist aber kein unterbrechungsfreier Zugriff möglich). Um diesen Flaschenhals zu entfernen, macht man HA (High Availability) mit mehreren Servern: Wenn ein Server ausfällt, übernimmt ein anderer unterbruchsfrei. Dadurch sind sogar Reserveserver nutzbar um den Zugriff im Normalbetrieb zu beschleunigen (Active-Active Konfiguration).

Darauf aufbauend können dann Dienste eingerichtet werden. Ein Beispiel aus vielen sind Geodaten:

- Damit Geodaten automatisch in einen Index aufgenommen werden können, in PostGIS, importiert und über Web Services bereitgestellt und schlussendlich mit R (via Rstudio) oder Python (via Jupyter) durch die Mitarbeitenden genutzt werden können, müssen diese zuerst strukturiert und organisiert werden. Das beinhaltet Management der Metadaten, garantierte Unveränderlichkeit der Rohdaten, nachvollziehbare Vorverarbeitung und Aufbereitung/Harmonisierung für einfache, schnelle Nutzung (z.B. Erstellen von Vorschaubildern für Rasterdaten «Pyramiden»). Diese Arbeiten benötigen sowohl GIS- als auch IT-Fachwissen.

Diese Werkzeuge und Services erlauben es, darauf aufbauen nachvollziehbare Analysen zu machen. Dazu gehören: Kenntnisse über die Datenquelle (z.B. garantiert unveränderte Satellitenbilder), nachvollziehbare Verarbeitung (z.B. automatisiert mit Skripten in R oder Python) und publizieren der Daten.

Begleitend muss laufend Dokumentation erstellt damit alle Mitarbeitenden wissen, welche IT-Ressourcen es gibt und wie sie diese nutzen können. [Git](#) ist zum Beispiel ein Werkzeug zur Versionsverwaltung um den Änderungen in einem Projekt zu verfolgen, ein sicheres Backup zu haben, mit Leuten gemeinsam an den gleichen Daten zu arbeiten und funktioniert problemlos offline. [Das könnte für die eine oder andere interessant sein](#).

# Übersicht Aufgaben

Hier ist eine Übersicht der aktuell anstehenden Arbeiten. Wir haben für jede Aufgabe den Aufwand geschätzt und dazu geschrieben, wer an der HAFL Interesse an der Aufgabe hat. Auf den weiteren Seiten sind die Aufgaben kurz aus technischer Sicht erklärt.

## Kategorie *Must*:

No	Task	Aufwand [Wochen]	Ansprechperson an HAFL
1	PostGIS	2	Alle GIS User
2	QGIS	2	Qgis User (v.a. Agronomie)
3	RStudio	1	WiMis
4	pgAdmin	1	Burgos
5	SciPy (numyp, pandas, etc.)	1	Burgos
6	NTW	1	Tiergenetiker
7	PhotoScan / WebODM	4	Berger, Burgos;
8	Maps / OpenLayers	2	Burgos, Spring
9	Geodata: Swisstopo	4	Alle GIS User
10	Geodata: Sentinel	1	Alle GIS User
11	Geodata: Landsat	1	Alle GIS User
12	Geodata: Modis	2	Waldwissenschaftler
13	Geodata: Webindex	2	Alle GIS User
14	Geodata: Webservice (OWS)	2	Alle GIS User
15	Jupyter	2	WiMis, Burgos
16	GPU rechnen	2	WiMis, Burgos
17	Reproducible science	on-demand	WiMis, Burgos
18	Paralles computing	on-demand	WiMis, Burgos
19	Samba HA	1	Alle GIS User
20	OpenData Kit	1	IL (Datenerhebung)
21	Intel MKL	1	WiMis

Total: 33 Wochen

## Kategorie *Nice to have*:

No	Task	Aufwand [Wochen]	Ansprechperson an HAFL
22	PXE	4	Alle GIS User
23	Terminal Server	4	Alle GIS User
24	Linux Images	4	Alle GIS User
25	Gitlab	4	WiMis, Burgos
26	Nextcloud	4	Alle
27	Share LaTeX	1	WiMis, Burgos, Kopp

Total: 21 Wochen

# Technische Beschreibung zu den Aufgaben

## 1 PostGIS: ermöglicht effizienten Zugriff auf zentrale Geodaten aus GIS Applikationen

- Mit [PostGIS](#) kann innerhalb der gängigen GIS Applikationen (QGIS, ArcGIS) via SQL direkt auf Vektor-Geodaten zugegriffen werden, ohne dass die Daten über den Umweg als Dateien via einen Fileshare geöffnet werden müssen. Dies macht Arbeiten mit Geodaten viel einfacher und effizienter, sowie beschleunigt Berechnungen/Analysen signifikant.
- **Tasks:** PostgreSQL Instanzen in Active-Active, PostGIS und Schema-Design und Implementierung, Import der Geodaten in SQL, Betrieb der Server, Dokumentation und Konfiguration Zugriff GIS-Applikationen.
- **Aufwand:** 4 Wochen

## 2 QGIS

- [QGIS](#) ist zusammen mit ArcGIS die am weltweit weitverbreiteste und an der BFH benutzte GIS-Applikation.
- **Tasks:** Aufrechterhalten der Pflege der Linux Paketierung, Betrieb der Server, Dokumentation, Beratung und Unterstützung der Nutzenden.
- **Aufwand:** 1-2 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 3 RStudio

- [RStudio](#) ist eine integrierte Entwicklungsumgebung für R (Statistik-Software). Mit RStudio Server ermöglichen wir auf unseren Number Cruncher für viele wissenschaftliche Mitarbeitende und Studierende die tägliche Arbeit mit RStudio via Browser.
- **Tasks:** Aufrechterhalten der Pflege der Linux Paketierung, Betrieb der Server, Dokumentation, Beratung und Unterstützung der Nutzenden.
- **Aufwand:** 1-2 Wochen, sowie je nach Bedarf für Beratung/Unterstützung (im 2017/2018 ca. 5% Job)

## 4 pgAdmin: Daten-Management in PostgreSQL-Datenbanken

- [pgAdmin](#) ist ein Webfrontend zum Management von Datenbanken in PostgreSQL.
- **Tasks:** Erstellen der Linux Paketierung, Installation und Betrieb der Server
- **Aufwand:** 1-2 Wochen

## 5 SciPy

- [SciPy](#) ist eine Sammlung von Programmbibliotheken und Werkzeugen für die Programmiersprache Python. Python wird immer mehr im wissenschaftlichen Bereich eingesetzt, unter anderem auch weil es verhältnismässig einfach zu lernen ist. Sehr viele Programme lassen sich mit Python automatisieren (z.B. ArcGIS, QGIS). ScyPy stellt funktionen für Datenorganisation, -verarbeitung, -analyse und -visualisierung bereit.
- **Tasks:** Aufrechterhalten der Pflege der Linux Paketierung.
- **Aufwand:** 1-2 Wochen

## 6 NTW

- Die Tiergenetik-Gruppe der HAFL verwenden spezielle Linux Software auf unseren Number Cruncher. Um diese weiterhin zu Benutzen zu können, sind dringend Upgrades notwendig.
- **Tasks:** Life Cycle von Debian 7 (EOL Q2/2018) auf Debian 9.
- **Aufwand:** 2 Wochen

## 7 PhotoScan / WebODM: Photogrammetrische Auswertungen (z.B. Drohnenbilder)

- [PhotoScan](#) und [WebODM](#) sind Programme welche 2D und 3D Bilder verarbeiten und analysieren koennen. An der HAFL werden so Drohnenbilder aufbereitet und ausgewertet. An der HKB werden im Rahmen der Konservation 3D Modelle verglichen.
- **Tasks:** Betrieb der Server, Erstellen der WebODM Paketierung, Beratung und Unterstützung der Nutzenden
- **Aufwand:** 4-6 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 8 Maps / OpenLayers: Visualisierung von Geoinformationen

- Mit [OpenLayers](#) können eigene «Google Maps» erstellt werden, welche gewünschte Informationen oder Zusammenhänge visualisieren.
- **Tasks:** Erstellen von gewünschten Visualisierungen
- **Aufwand:** 1-2 Wochen für bereits gewünschtes, weiteres nach Bedarf

## 9 Geodata: Swisstopo

- An der BFH werden an diversen Orten Geodaten der Swisstopo abgelegt. Diese müssen zentral gesammelt, konsistent abgelegt und mit Metadaten (u.a. Lizenzierung) versehen werden.
- **Tasks:** Neubeschaffung bestehender/veralteter/unvollständiger Swisstopo Datensätze, Erstellen der Metadaten und Formate zur weiteren Verarbeitung
- **Aufwand:** 8-12 Wochen

## 10 Geodata: Sentinel

- Daten der Sentinel-Satelliten werden an der BFH für einige Projekte verwendet.
- **Tasks:** Aufrechterhalten des bestehenden täglichen Imports von Sentinel 2 (Band 2), Ablage und Konvertierung der Satellitendaten, Betrieb der Server, Beratung und Unterstützung der Nutzenden, Erweiterung auf Sentinel 2 (Band 1) und weitere Sentinel Satelliten (Sentinel 1-7)
- **Aufwand:** 2-4 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 11 Geodata: Landsat

- **Tasks:** Aufbau Landsat Satellitendaten analog Sentinel
- **Aufwand:** 2-4 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 12 Geodata: Modis

- **Tasks:** Aufbau Modis Satellitendaten analog Sentinel
- **Aufwand:** 2-4 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 13 Geodata: Webindex

- Damit die Nutzenden einfach prüfen können, welche Geodaten in welcher Qualität/Grösse/Format/Aktualität/Lizenz an der BFH vorhanden sind, wird ein durchsuchbarer Webkatalog aufgebaut.
- **Tasks:** Aufbau Webkatalog basierend auf zentral abgelegten Geodaten und deren Metadaten, ggf. in Zusammenarbeit mit externer Firma
- **Aufwand:** 2-4 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 14 Geodata: Webservice (OWS)

- Analog zu Vector-Geodaten welche mit PostGIS bereitgestellt werden müssen, können Rasterdaten («Fotos») mit entsprechender Software als Dienst zur Verfügung gestellt werden. Damit werden Webdienste ähnlich zu «Google Maps» mit den eigenen Resultaten möglich.
- **Tasks:** Evaluation geeigneter Software, Aufbau und Betrieb des Dienstes, Dokumentation und Konfiguration Zugriff GIS-Applikationen.
- **Aufwand:** 1-2 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 15 Jupyter

- [Jupyter](#) ist eine integrierte Entwicklungsumgebung für viele gängige Sprachen/Software (R, Octave, LaTeX, Python und vieles mehr). Damit können Studierende wie auch wissenschaftliche Mitarbeitende unsere Number Cruncher nicht nur bequem uebers Web mit R benutzen, sondern für die Mehrheit der wissenschaftlichen Arbeit.
- **Tasks:** Erstellen der Linux Paketierung, Betrieb der Server, Unterstützung/Beratung der Nutzenden
- **Aufwand:** 1-2 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 16 GPU Rechnen

- Unsere Linux Number Cruncher haben spezielle GPU Rechenkarten mit welcher alle bisherigen wissenschaftliche Berechnungen um mehrere Größenordnungen schneller werden.
- **Tasks:** Anpassungen in der Linux Container Toolchain, Unterstützung/Beratung der Nutzenden
- **Aufwand:** 1-2 Wochen, sowie je nach Bedarf für Beratung/Unterstützung

## 17 Reproducible Science

- In der Wissenschaft müssen Resultate nachvollziehbar sein. Dazu gehört immer mehr, dass auch die verwendeten Daten publiziert werden. Damit publizierte Daten auch für andere sinnvoll sind, müssen sie aufbereitet werden. Die Arbeit des Aufbereitens erst am Schluss des wissenschaftlichen Prozesses anzuhängen macht aber wenig Sinn, denn grundsätzlich profitiert jede Arbeit von nachvollziehbarer Datenverarbeitung und reproduzierbarer Analyse.
- **Tasks:** Hier ist vor allem die Unterstützung mit Werkzeugen und Infrastruktur im Hinblick auf Reproducible Science gemeint.
- **Aufwand:** Je nach Bedarf.

## 18 Parallel Computing

- Computer Prozessoren werden nicht mehr wesentlich schneller, dafür haben wir immer mehr davon. Geschwindigkeitssteigerungen kann man nur noch erreichen, wenn man Aufgaben Parallel rechnet, d.h. mehrere Prozessoren rechnen an der gleichen Aufgabe. Damit Parallelisierung zuverlässig funktioniert, müssen die Prozeduren und Programme angepasst werden
- **Tasks:** Unterstützung in der Auswahl der richtigen Tool und Methoden um parallel zu rechnen
- **Aufwand:** Je nach Bedarf.

## 19 Samba HA

- Unserem Ceph Storage Cluster für Big Data ist ein Samba Gateway vorgeschaltet, damit auch nicht-Linux Nutzende bequem über SMB Zugriff auf Ceph haben. Momentan steht nur ein einziges Samba Gateway zur Verfügung, welches aus Zuverlässigkeitsgründen unbedingt redundant ausgelegt werden muss.
- **Tasks:** Setup Samba High-Availability
- **Aufwand:** 2 Wochen

## 20 OpenData Kit

- [OpenDataKit](#) ist ein System um mit Mobiltelefonen (oder Tablets) Datenerhebungen im Feld zu machen. Es gibt ein App für Smartphones, dass man im Feld offline nutzen kann. Den Server (sozusagen die Cloud dafür) muss man selber betreiben.
- **Aufwand:** 1-2 Wochen

## 21 Intel Math Kernel Library

- Intel stellt (neuerdings kostenlos) optimierte Varianten von u.a. BLAS/LAPACK zur Verfügung. Damit werden Matrizen-Operationen in R auf unseren Number Crunchern gemessen(!) 2x so schnell. Die Intel MKL muss paketiert und verteilt werden.
- **Aufwand:** 1-2 Wochen

## 22 PXE

- An der BFH steht ein veraltetes Linux Live System über Netboot (PXE) zur Verfügung. So kann jeder Rechner an der BFH durch eine Auswahl beim Start in eine Linux Umgebung gebootet werden. Dies ist insbesondere für Lehre angenehm, da so BYOD genutzt werden kann, ohne dass Software installiert werden muss.
- Das veraltete System muss neu erstellt werden und mit aktueller, GIS relevanter Software ergänzt werden.
- **Aufwand:** 4 Wochen

## 23 Terminal Server

- Damit Mitarbeitende ohne Reboot ihres Arbeitsplatzrechners Linux Applikationen verwenden können, sollen diese via VNC, RDP und HTML5 gestartet werden (etwa vergleichbar zu Citrix). Dafür muss ein Terminal Server aufgebaut werden.
- **Aufwand:** 4 Wochen

## 24 Linux Images

- Damit unsere Linux Umgebung nicht nur via Netzwerk sondern auch offline/zuhause etc. benutzt werden kann, stellt die ITS kombinierte Installer und Live Images zur Verfügung.
- Die bestehenden Images sind veraltet, müssen aktualisiert und um entsprechende GIS Software ergänzt werden.
- **Aufwand:** 4 Wochen

## 25 Gitlab

- Die ITS stellen ihren eigenen Git-Server der ganzen BFH zur Verfügung. Dessen Bedienung erfüllt die Anforderungen der ITS intern, nicht aber aktuelle Kundenwünsche.
- [Gitlab](#) stellt eine angenehmer, zeitgemässer Self-Service für das Management von Git Repositories zur Verfügung. Dies ist elementar für die breite Anwendung von Git.
- **Aufwand:** 4 Wochen



## 26 Nextcloud

- Momentan existiert an der BFH keine plattformunabhängige Möglichkeit, eigene Dateien von überallher zur Verfügung zu stellen. Insbesondere gibt es keine Möglichkeit, grosse Dateien mit externen Partnern austauschen zu können.
- [Nextcloud](#) (wird u.a. auch für SWITCHdrive verwendet) kann mit unserem Ceph Storage Cluster integriert werden, so dass die Daten als «Cloud» verfügbar sind.
- **Aufwand:** 4 Wochen

## 27 ShareLaTeX

- [ShareLaTeX](#) ist ein einfach zu bedienender Online-Editor für LaTeX-Dokumente. Gleichzeitiges Zusammenarbeiten und Integration in Version Control sind ebenso möglich.
- **Aufwand:** 1 Woche

Letzte Aktualisierung: 10. April 2018

Simon Spöhel, Daniel Baumann