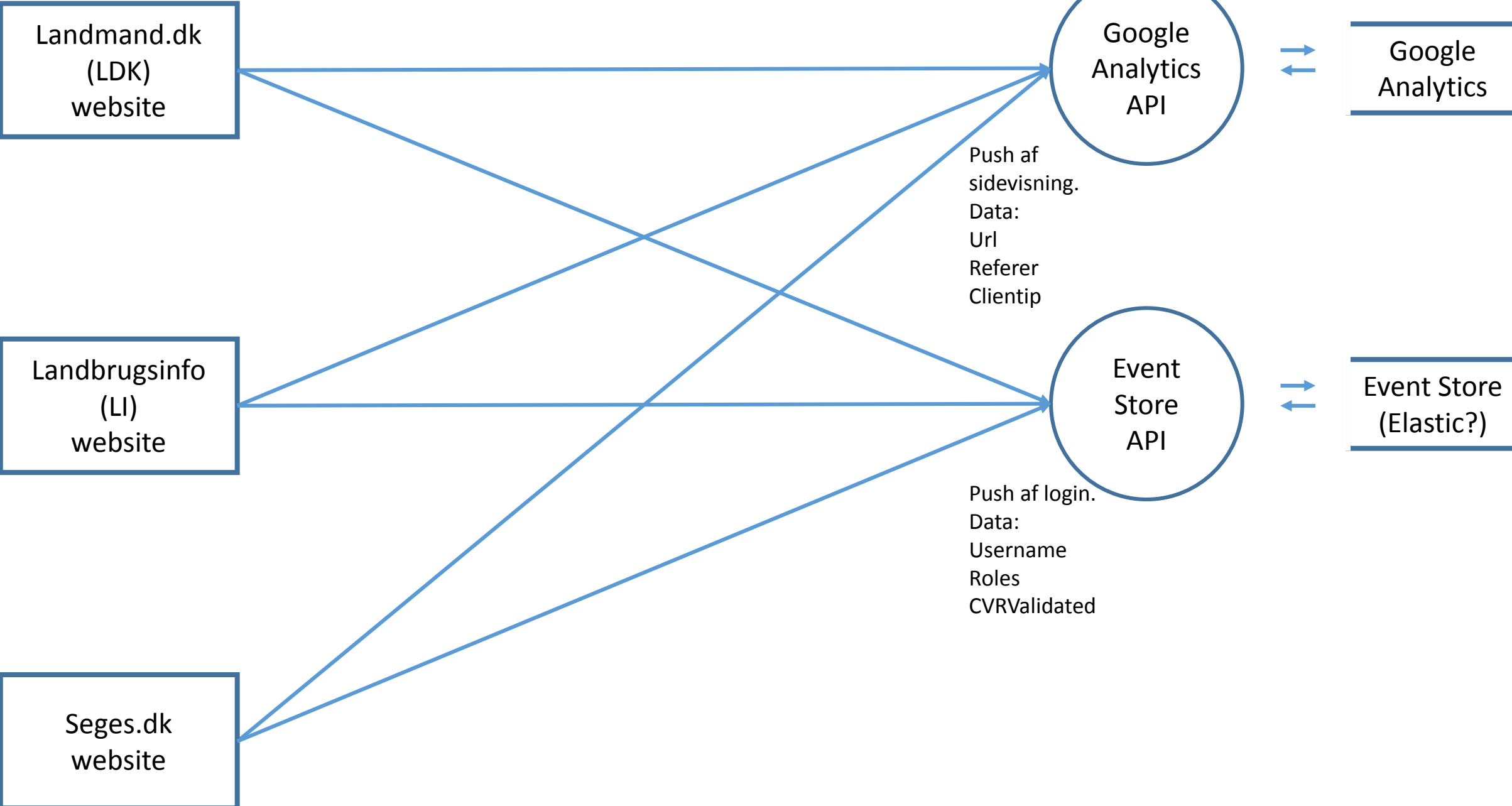


ElasticSearch / Kibana forbehold

- ElasticSearch er et datastore, der kan opbevare og søge effektivt i store, semi-strukturerede datamængder
- Kibana er et web-baseret præsentationsværktøj, der kan visualisere data fra ElasticSearch
- SEGES installation af ElasticSearch/Kibana foretager ikke authentication/authorization (authn/authz) af brugere
 - alle data der gemmes er således tilgængelige for enhver PC tilsluttet SEGES netværk
- Der gemmes i dag store mængder persondata fra andre løsninger i ElasticSearch
 - det må forventes nogen en dag får gennemtruffet at authn/authz er nødvendigt
 - den dag vil løsninger der anvender ElasticSearch/Kibana skulle efter-implementere authn/authz.
- Projektet skal kunne bære denne risiko
 - den økonomiske for, at der på et uforudsigeligt tidspunkt skal bruges penge på at implementere authn/authz
 - den juridiske, såfremt der gemmes persondata i ElasticSearch af denne løsning

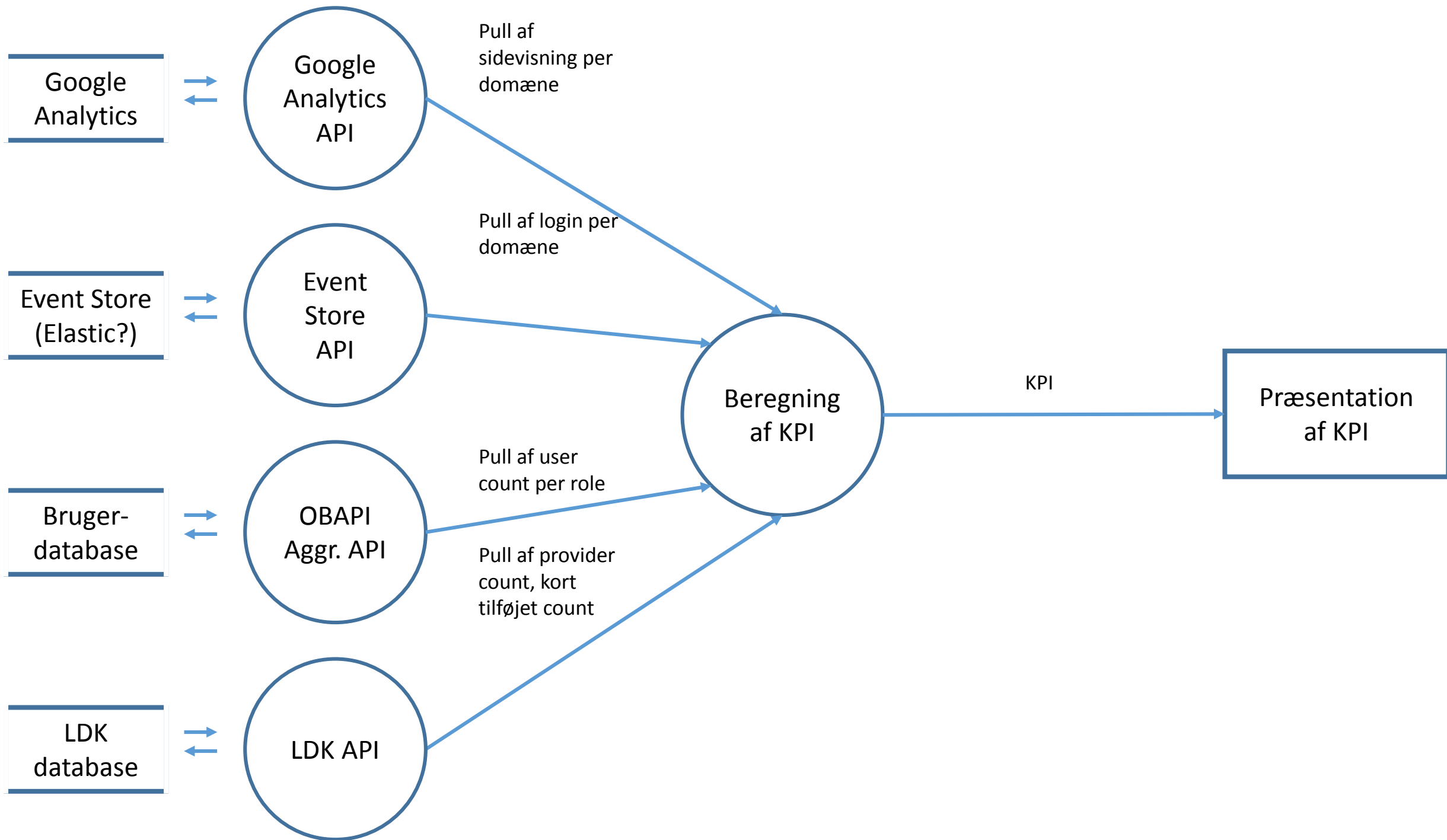
Dataopsamling

- Google Analytics opsamler i dag de nødvendige data om sidevisninger på de tre platforme
- Loginhændelser opsamles ikke
 - Det skal opsamles i et event store, eksempelvis Serilog/Elasticsearch logging i de respektive systemer



Beregning af KPI'er – nødvendigt data

- Sidevisninger (fra Google Analytics)
 - Clientip
 - Det skal afklares om gruppering af clientip -> ekstern/intern trafik opsættes i GA, eller som en del af beregningsmotoren
- Loginevents med username/roles (fra Event Store)
 - Aktive medlemmer
 - Aktive medlemmer og CVR-validerede
 - DLBR-ansatte
- Aggregerede data om alle brugere (fra OBAPI)
 - Antal aktive medlemmer
 - Antal DLBR ansatte
 - Aggregeringen bygges som nye endpoints, baseret på de gamle politikker
 - Må man forespørge brugere indenfor en given gruppe, så må man også bede om antallet af brugere i samme
- Kort og providerdata (fra Landmand.dk)
 - Antal kort tilføjet via "vis og skjul services", aggregeret som eet tal for alle brugere (korrekt?)
 - Antal portalserviceejere med mindst én portalservice tilknyttet (korrekt?)



Abstrakt datamodel -> Konkret teknologi

- Beregning af KPI, samt præsentation af disse kan løftes af en række forskellige teknologier
 - Præsentationen skal være web-baseret
 - Authentication er nice-to-have, men ikke strengt nødvendigt for de aktuelle KPI'er, der aggregerer personfølsomme data væk
- En hjemmestrikket løsning kunne bestå af
 - C# batchjob (task scheduler eller Hangfire) afvikles dagligt
 - trækker data
 - beregner KPI
 - pusher KPI-records til Elasticsearch
 - Kibana
 - Visualiserer KPI-records på dashboard
- Alternativer baseret på PowerBI, Reporting Services eller andet er også en mulighed
 - Man kan håbe på at noget af det vi ellers selv skal kode er OOTB
 - Vi (kælderteamet) har ikke megen erfaring med disse teknologier
 - Vi bør derfor trække på husets kompetencer (Kvæg, Plante, Økonomi), hvis vi går ned af den sti

Opgaver (I)

- Udvid OBAPI med endpoint, der kan levere count(*) aggregering af antal brugere medlem af en eller flere grupper
 - Skal respektere den for den forespørgende systembruger opsatte politik, dvs. hvilke grupper vedkommende må spørge på
- Udvid LI, LDK og SEGES.DK med Serilog-logning af loginevents, inklusiv CVR-valideringsstatus og roles på brugeren
 - Brugerid bør anonymiseres at rest (fx som peppered hash), vel vidende at anonymiseringen vil være trivielt at tilbageføre med brute force for os (da vi kender pepper)
 - Ikke-nødvendige gruppemedlemskaber bør smides væk inden logning

Opgaver (II)

- Batch-kørsel der kan trække på datakilderne og producere KPI'er, der gemmes som nye events
- Kibana dashboard, der viser KPI-events i passende visualisering