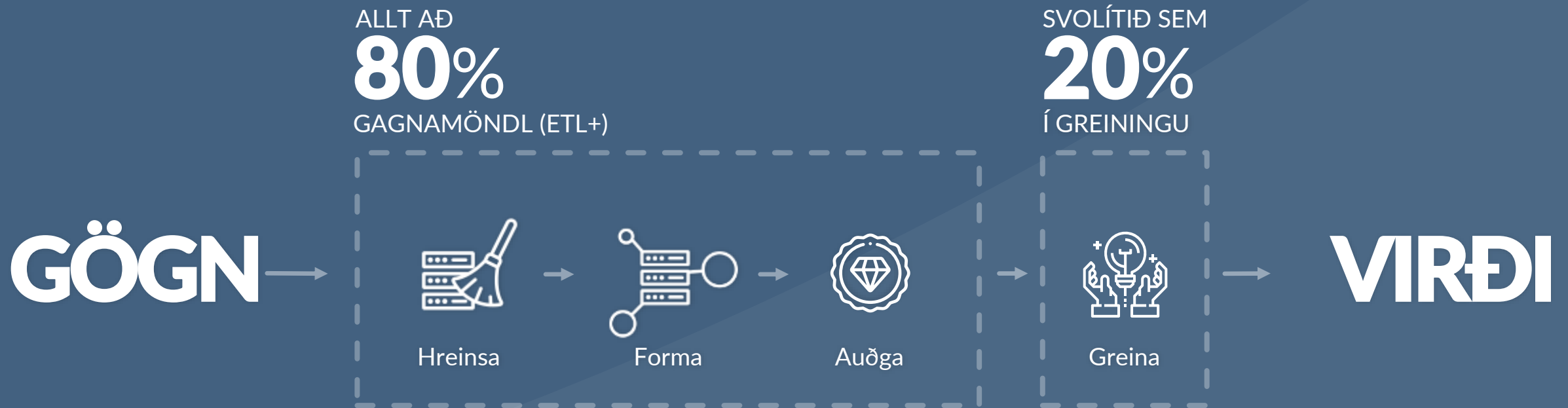




Snjallgögn

Fyrirvara- og fyrirhafnarlaus hagnýting opinna gagna

Ótrúleg (og síendurtekin) tímasóun



Langa leiðin ... frá opnum gögnum að hagnýtingu



Aðgengi

Hagnýting

Langa leiðin ... frá opnum gögnum að hagnýtingu



Skilmálar
(Hamlandi)

Aðgengi



Hagnýting



Snjallgögn

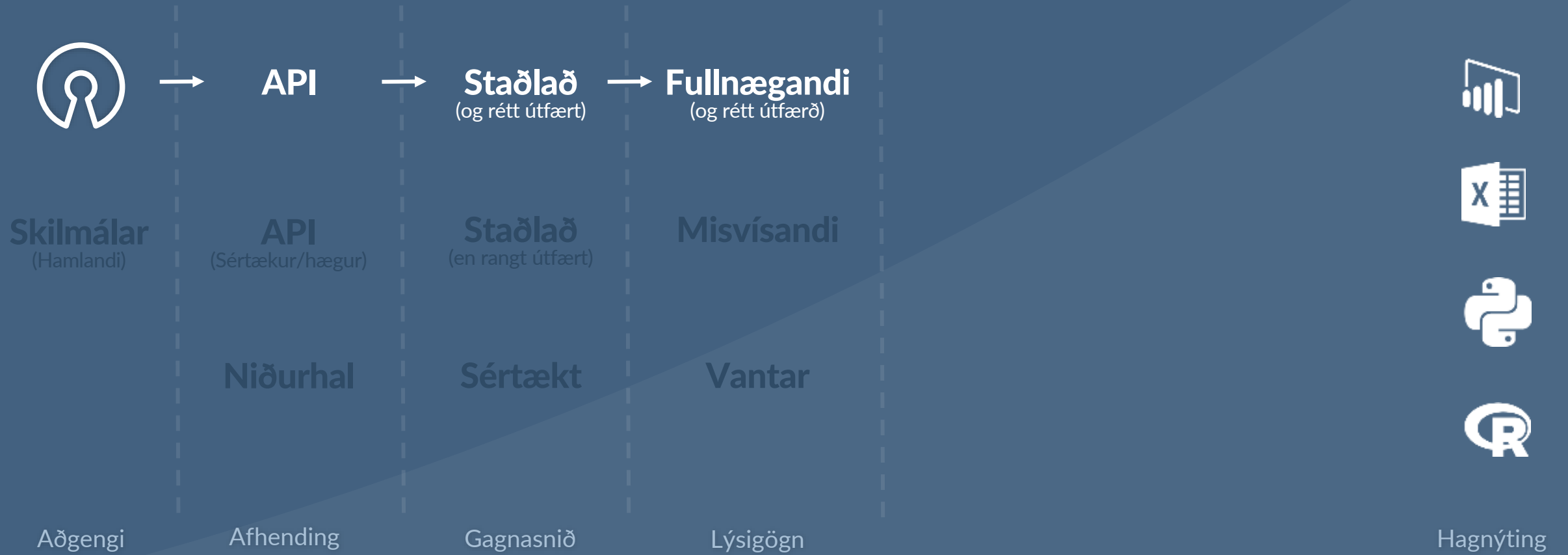
Langa leiðin ... frá opnum gögnum að hagnýtingu



Langa leiðin ... frá opnum gögnum að hagnýtingu



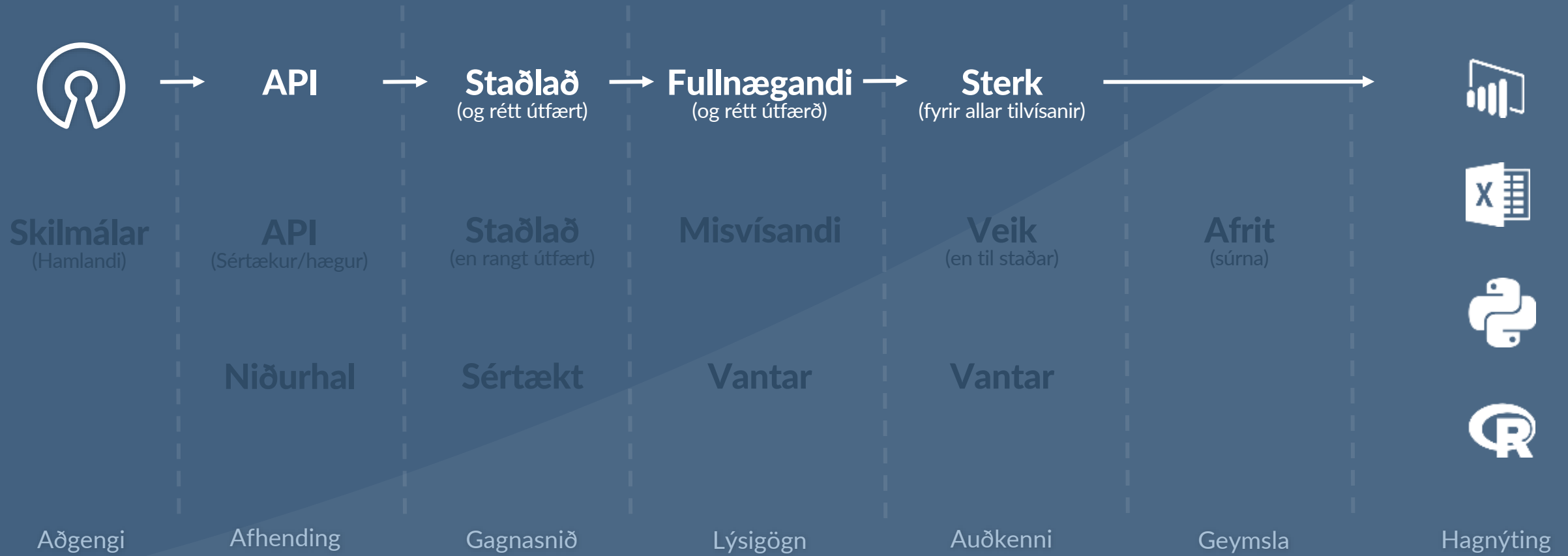
Langa leiðin ... frá opnum gögnum að hagnýtingu



Langa leiðin ... frá opnum gögnum að hagnýtingu



Langa leiðin ... frá opnum gögnum að hagnýtingu



Hvar endar ábyrgð/áhugi útgefenda?



API

ætlaður fyrir
rauntímavinnslu



Staðlað

og rétt útfært
gagnasnið



Fullnægandi

lýsigögn fyrir öll gildi,
mælieiningar og kerfi



Sterk

og stöðluð auðkenni
fyrir allar tilvísanir



- Opnun gagna er ekki hið endanlega markmið
- hagnýting þeirra er það!
- Eini alvöru mælikvarðinn á hversu vel hefur tekist til
... er hvort að það er hægt að nýta þau með fyrirhafnar- og fyrirvaralausum hætti
- Tökum ábyrgða "alla leið"



Gott að hafa í huga



- Ekki nota íþyngjandi skilmála nema í neyð
 - Ef markmiðið er að hámarka virði þá þurfa notendur að geta nýtt gögnin



Gott að hafa í huga



- Léttir staðlar til sem henta fyrir jafnt lítið og mikið magna gagna
 - Hýsing í skýinu leyfir sjálfvirkt viðbragð við álagi og lágmarkar kostnað utan álagstíma
- API aðgangur hannaður fyrir streymandi upplýsingar sem og söguleg gögn
- Besti API-inn er nýtanlegur beint úr endabúnaði notenda
- Lágmarka þarf svartíma fyrir beina notkun
 - frekar en afhendingu
 - það má kosta að fá aðgang að "rekstrar API" frekar en aðgang að "niðrhals API"



Gott að hafa í huga



- Vond útfærsla á stöðluðu gagnasniði getur verið verri en (gott) sérsnið
 - ISO 8601 fyrir dagsetningar, tímabil o.fl.
- Notum alþjóðlega og viðurkennda staðla fyrir tölur, dagsetningar o.s.frv.
- CSV getur td. verið 100% en það getur líka verið svo vont
- Reynum að forðast gagnasnið með mikla yfirbyggingu
 - XML og JSON eru ekki æðisleg snið fyrir mikið magn endurtekinna gagna
 - Það eru til binary snið sem styðja vef og "evolving-schema" með góðum hætti (Avro, Protobuf o.fl.)
- JSON-Stat er dæmi um staðal fyrir töluleg gögn
 - Aðallega notað af stærri alþjóðlegum stofnunum



Gott að hafa í huga



- Góð lýsigögn tryggja rétta meðhöndlun og hjálpa til við vörpun
- Mælieiningar fyrir allar tölulegar upplýsingar (Unit of Measure)
 - Fyrir hvað stendur talan og á hvaða skala er hún?
 - UN CEFACT er gott dæmi um samræmdar upplýsingar um einingar
- Upphæðir (sem dæmi)
 - Gjaldmiðill, uppreiknað gildi?, dagsetning gjaldmiðilsstöðu
- Ekki setja lýsigögn í dálkanöfn
 - Notið hliðardálka fyrir breytileg gildi
 - Það er auðveldara að nýta lýsigögn sem eru samofin gögnunum en lýsigögn í skjölun



Gott að hafa í huga



- **Góð gögn innihalda sterk auðkenni fyrir allar tilvísanir**
 - Nafn á landi er t.d. veikt auðkenni.
 - Það eru til ótal staðlar til að auðkenna nánast hvað sem er
- **Tengd gögn verða til með sterkum tilvísunum**
 - 5 stjarna Tim Berners-Lee (SW) fæst þegar sterka vísanir eru í samhengisupplýsingar (context)
 - Hinn merkingarbæri vefur byggir á þessari hugmynd
- **Innri tilvísananúmer skaffa innri tengingar**
 - Sterkar tilvísanir hjálpa til við allar ytri tengingar



Gott að hafa í huga



- Stótt gögn súrna
- Þegar gagnamöndl er nauðsynlegt er geymsla oft eini valmöguleikinn
- Það er merki um lélega útfærslu þegar þess er þörf
- Notendur sem ráða ekki yfir geymsluinnviðum eiga að getað notað gögnin beint úr þeim hugbúnaði sem þeir nota til greininga





Snjallgögn