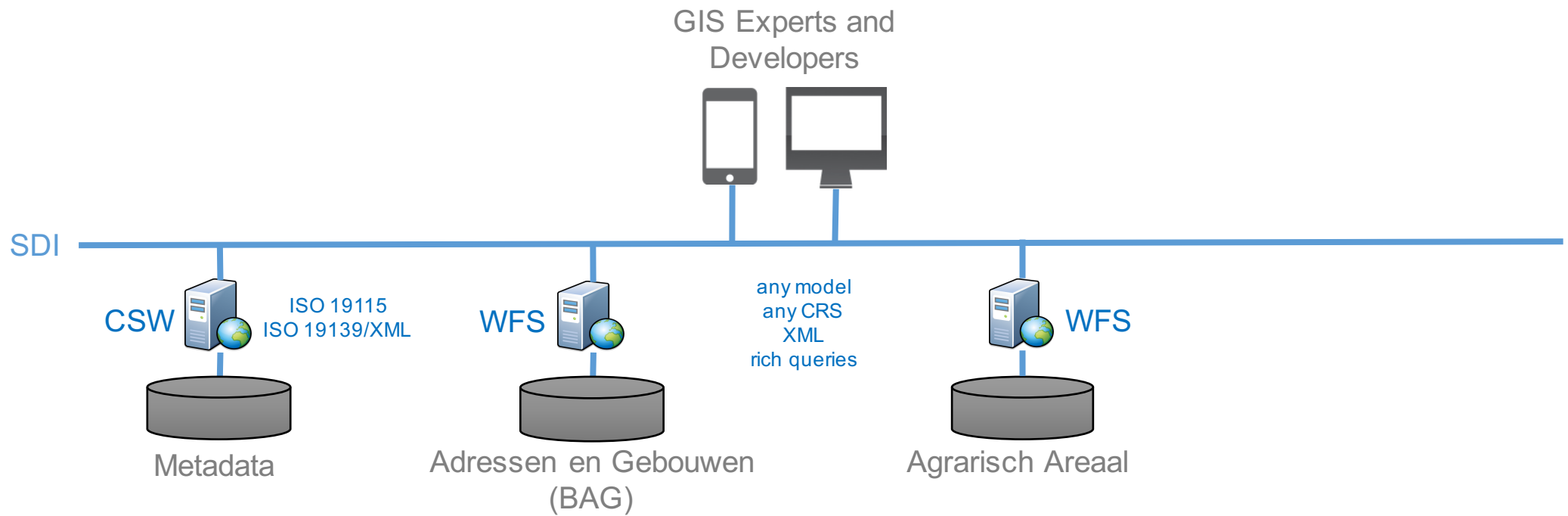


Geonovum Testbed – Topic 4  
“Spatial data on the Web using the current SDI”

**Final Event**



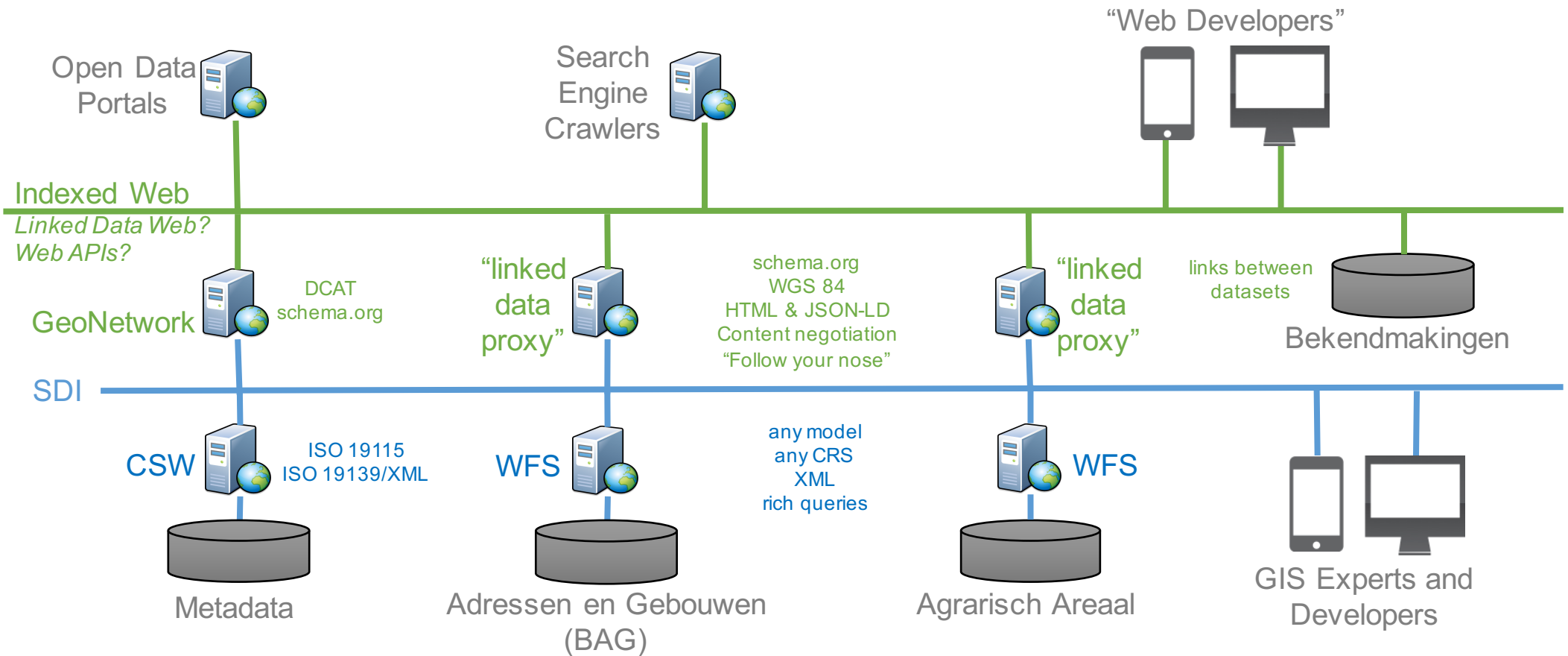
# The starting point ...



# What are we trying to do?

- crawlability and linkability, i.e. make each resource hosted by a WFS or CSW available via a persistent URI and ensure that all resources can be reached via links from a “landing page” for a data set ( = **presence on the Web of data**)
- classification of the resources using vocabularies supported by the main search engines on the Web ( = **harmonisation of data discovery**)
- representations of data for consumption by humans (HTML), developers (JSON-LD, GeoJSON, GML) and search engine crawlers (HTML with structured data annotations) ( = **data access based on current Web practices**)
- establishing and maintaining links between data ( = **connecting data with other data on the Web**)
- discovery of both spatial and non-spatial data by the same search engine ( = **harmonisation of data discovery**)

# What have we built



# Demo overview

- Spatial Data (via a WFS proxy)
  - Clemens Portele, interactive instruments GmbH
- Metadata (via a CSW proxy)
  - Paul van Genuchten, GeoCat BV
- Linking and Linked Data
  - Lieke Verhelst, Linked Data Factory

# Architectural principles used in the proxy design

- All resources are identified using persistent HTTP URIs
- All resources are discoverable via search engines
- All interaction is using the HTTP protocol and consistent with its design
- APIs to access data should be self-describing and support immediate use
- Resources can be accessed and understood by developers and citizens
- Resources are either explicitly linked using HTTP URIs or data is structured so that links can be established dynamically

# Summary of the Findings

- To a large extent we were successful → see demos
- But there is still a considerable way to go, e.g.
  - Data integration across datasets: Properly structured data using common vocabularies will decrease the cost of data integration (here: linking) significantly
  - Metadata in WFS capabilities and the dataset metadata in Catalogs is often incomplete or inconsistent
  - Performance & data compactness: Response times and data size → impact on the indexing / ranking by search engines and usability in general
  - Search engines are largely a black box when we look at spatial data: how is structured spatial data used?
  - schema.org is a divergent kind of vocabulary - from the Linked Data perspective
  - Different vocabularies needed for different use cases / communities – e.g. schema.org vs GeoSPARQL
  - Content negotiation based on media types does not support different vocabularies / user communities

# Report (70+ pages)

- <https://github.com/geo4web-testbed/topic4-general/blob/master/spatial-data-on-the-web-using-sdi-report.pdf>
  - for now as PDF, HTML to come (Google Docs to HTML does not look good enough...)
- Resources that are made available by the proxies
- Our strategy for assigning URIs to the resources
- The mapping of the data and metadata resources to the schema.org vocabulary (and GeoDCAT-AP)
- The representations (formats), in which each resource is available
- Our experiences with establishing links across datasets and between data and metadata
- Our experiences with search engines crawling and indexing the resources
- Details about our findings and open challenges
- Do our proxies implement the current draft W3C/OGC best practices for spatial data on the Web?