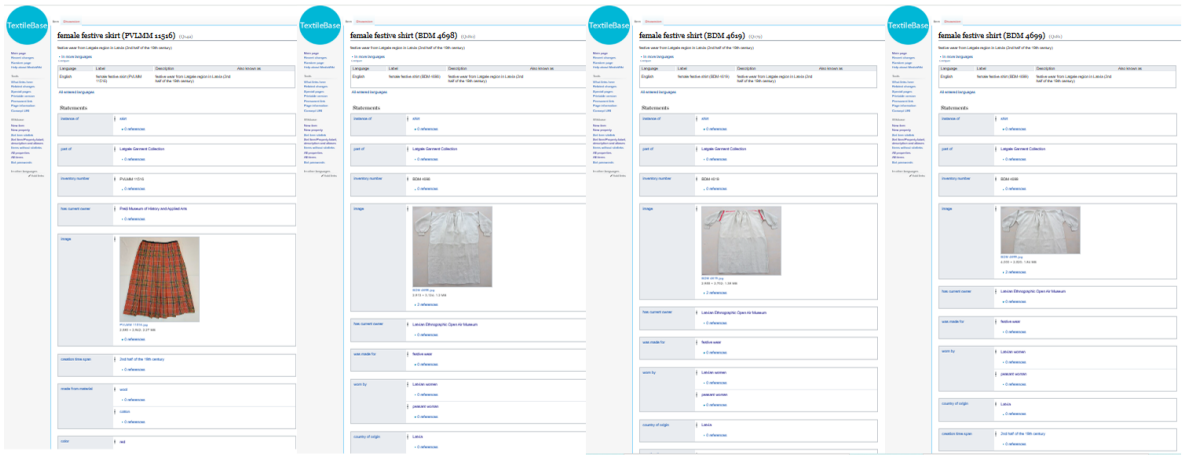
Reprex B.V.
Saturnusstrat 14, 2516 AH
's-Gravenhage, the Netherlands

TextileBase — Technical Specifications

`TextileBase` is a platform to **collect, connect, and share research and metadata on historical clothing**. It links artefacts, photographs, secondary sources, and institutional records into a multilingual, interoperable, and searchable knowledge graph.

Before subscribing, you may want to learn more about the project itself:



TextileBase overview ▶ TextileBase main page ▶ TextileBase methodology preprint ▶ TextileBase subscription offerings ▶ TextileBase seminar (with slides)

1) System Architecture

- **Core stack:** Wikibase (MediaWiki + WB stack), MariaDB/PostgreSQL, RDF triplestore (Apache Jena Fuseki (optional: Blazegraph or Virtuoso)), SPARQL 1.1 endpoint.
- **Knowledge graph services:** RDF export, URI resolver, PID minting integrations (DOI for artefacts, datasets, data papers, ISCC for their file manifestations).
- **Discovery UI:** Sampo-style UI for faceted search and one-click queries; optional embedded widgets for partner sites.
- **Pipelines:** ETL jobs for import/cleaning (CSV/TSV, JSON, XML, IIIF), scheduled reconciliation, export jobs (RDF/CSV/JSON/ZIP).
- **APIs:** MediaWiki/Wikibase APIs, SPARQL endpoint, REST export endpoints, webhook notifications for change events.
- **Deployment:** Containerized (Docker), orchestrated (Docker Compose/Kubernetes), staging + production environments, IaC-ready.

Reprex B.V.
Saturnusstrat 14, 2516 AH
's-Gravenhage, the Netherlands

2) Data Model & Semantics

- **Ontologies:** CIDOC-CRM (museum/heritage), RiC (archives), DCTERMS (generic semantics), SKOS (thesauri), PROV-O (provenance). We use ontological patterns to enable the use of the simple, citizen scientist or non-technical researcher graphical user interface of Wikibase. The patterns expressed in the Wikibase Data Model are translated to standard ontologies.
- **Entity types:** Artefact, Digital surrogate, Person/Agent, Place, Collection, Institution, Concept (vocabularies), Event/Activity, Work/Publication.
- **Relations:** Authorship/Creation, Collection membership, Provenance/Ownership history, Fabrication technique/material, Depictions, Citations/References.
- **Constraints & validation:** Property constraints (cardinality, allowed values via SKOS schemes), SHACL/SHEx checks in curation pipelines.

3) Identifiers & Provenance

- **Persistent IDs:**
  - *External PIDs:* DOI (DataCite), ISCC for content fingerprints, ORCID for contributors, ROR for institutions, VIAF/ISNI for authority links.
  - *Internal:* Stable HTTP URIs for all items; slugged + numeric IDs with redirects on merge.
- **Provenance & versioning:** Full edit history (Wikibase), change logs for ETL, PROV-O statements for source and transformation lineage.

4) Multilinguality & Thesauri

- **Labels/descriptions:** Multilingual labels and aliases; language fallback rules.
- **Authority links:** Wikidata, Getty AAT/TGN/ULAN, Geonames, Library of Congress, National authority files.
- **Lexical support:** Normalization for historical spellings; synonym/variant tables; lemmatization for search (per language). Lemmatization is used in combination with premium AI services.

5) Data Ingest & Reconciliation

- **Inputs:** Excel/CSV/TSV, JSON, XML (EAD/METS/TEI subset), IIIF manifests, OAI-PMH harvests.
- **Tools:** OpenRefine (with Wikibase extension), scripted ETL (Python/R), bulk uploaders, reconciliation against external authorities.
- **Media:** Image attachment via IIIF URLs or file store; checksum verification; optional derivative generation (thumbnails, tiles).
- **Quality checks:** Schema conformance, controlled vocabulary enforcement, date/place parsing, duplication detection.

Reprex B.V.
Saturnusstrat 14, 2516 AH
's-Gravenhage, the Netherlands

6) Search, Query & Access

- **Text search:** Full-text with facets (language, time, place, type), autocomplete, typo tolerance.
- **Graph queries:** SPARQL 1.1 endpoint; example:

```
  SELECT ?item ?itemLabel ?placeLabel ?date WHERE {
    ?item wdt:P31 wd:Q_Artefact ;
          wdt:P131 ?place ;
          wdt:P571 ?date .
    FILTER(?date >= "1800-01-01"^^xsd:date && ?date < "1900-01-01"^^xsd:date)
    SERVICE wikibase:label { bd:serviceParam wikibase:language "en,et,lv,fi,h
u" }
  }
  LIMIT 50
```

- **Exports:** RDF (Turtle/JSON-LD/N-Triples), CSV/TSV, JSON dumps; dataset ZIPs with metadata and README.
- **Programmatic access:** REST export routes, SPARQL, MediaWiki/Wikibase APIs, IIIF links.
  Note: Programmatic access is available only for Premium packages.

7) AI-Assisted Modules

- **Core (Starter & above):**
  - Multilingual search expansion
  - Historical place-name resolution
- **Extended (Standard & Premium):**
  - Synonym/variant matching
- **Advanced (Premium only):**
  - Image/text metadata extraction (OCR / handwritten OCR where supported)
  - Visual similarity search and "possible match" alerts (confidence threshold adjustable)
  - Semi-automatic entity linking with curator-in-the-loop review
  - Lemmatization
- **Explainability:**
  - All AI suggestions are logged with confidence scores
  - Curator approval is required before publication

Reprex B.V.
Saturnusstrat 14, 2516 AH
's-Gravenhage, the Netherlands

8) Research Data Management (RDM)

- **FAIR/8-star workflows:** metadata completeness checks, PID issuance, machine-readable licenses, landing pages, citability
- **Repository integration:** Zenodo (DOI minting), EU Open Data Portal, institutional repositories; scripted crosswalks
- **Data papers:** templated documentation; reproducible SQLite/RDF bundles; citation snippets (APA/Chicago/MLA)
- **Reporting:** periodic KPIs (records added, PID coverage, linkouts, reuse metrics when available)

9) Hosting, Security & Compliance

- **Availability:** 99.5–99.9% SLA target (tier-dependent), CDN for static assets
- **Backups:** automated daily DB backups, weekly off-site snapshots, test restores; versioned object storage for media
- **Security:** HTTPS/TLS, role-based access control (RBAC), 2FA for admins, audit logs, least-privilege service accounts
- **Privacy & GDPR:** data minimization; DPA on request; PII handling policies; configurable retention; consent records for images where applicable
- **Licensing:** open data defaults (CC0/CC BY) configurable per partner; embargo support

10) Sampo UI & Embedding

- **Faceted discovery:** time/place/type facets; saved queries; permalinked visualizations
- **Visualizations:** maps (place of origin/provenance), timelines, network graphs; export as PNG/SVG/CSV
- **Embedding:** widgets/iframes for partner sites; query URLs for live embeds

11) Digital Passport (Business)

- **Profile:** per-item passport page with materials, origin, production method, heritage links, sustainability fields
- **Identifiers:** DOI/ISCC, QR code generator for labels; resolvable URLs
- **Schema:** JSON-LD product metadata aligned with EU DPP concepts (materials, durability, repairability)

Reprex B.V.
Saturnusstrat 14, 2516 AH
's-Gravenhage, the Netherlands

- **Integration:** JS snippet or server-side include for webshop; webhook to update on catalogue changes
- **Optional analytics:** scan counts, referrals, geo (aggregate/anonymized)

12) Tier Limits & Options (defaults; adjustable in contract)

- **Starter:**
  - Records: 200–500; one curated dataset (≤1,000 rows)
  - No programmatic API or SPARQL access (exports only: CSV, RDF, SQLite)
  - PID minting: up to 20 per year (DOI/ISCC/URI)
  - Light AI only: multilingual search expansion + place-name resolution
  - Storage: 5–10 GB media; 1 scheduled export/month
  - Support: email only (≤5 business days response); onboarding checklist; DMP template pack
- **Standard:**
  - Records: ≤20,000 across ≤5 datasets
  - Programmatic access: limited SPARQL + API (priority queue)
  - PID minting: expanded (DOIs/ISCCs/URIs as needed)
  - AI: includes synonym/variant matching
  - Storage: ≤200 GB media; weekly exports; hosted collection section
  - Support: email + video; 1 staff workshop; quarterly QA review
- **Premium:**
  - Records: unlimited (fair-use); all AI modules enabled
  - Programmatic access: dedicated API/SPARQL capacity; SLA-backed; staging environment
  - PID minting: unlimited (as needed for consortium scale)
  - Storage: multi-TB; nightly exports; custom dashboards
  - Support: priority line; annual onsite/hybrid workshop; custom ontology & thesaurus building

13) Performance Targets

- **Entity create/edit latency:** < 1s UI commit; < 5 min KG sync
- **Search:** < 500 ms (95th percentile) for common queries at nominal load
- **SPARQL:** < 2 s (95th percentile) for indexed patterns; guidance for heavy joins

14) Change Management & Interop

- **Schema evolution:** versioned property sets; migration scripts; deprecation notes
- **Crosswalks:** mappings to Europeana EDM, Dublin Core, MARC-lite/CSV; IIIF harvesting

Reprex B.V.
Saturnusstrat 14, 2516 AH
's-Gravenhage, the Netherlands

- **Partner integration:** OAI-PMH (optional), CSV dropbox/HTTPS endpoints, Git-based config repos

15) Documentation & Training

- **Docs:** contributor guide, data model handbook, API cookbook, SPARQL recipes
- **Templates:** DMPs, data paper skeletons, grant text blocks (FAIR/impact)
- **Training:** Starter webinar (1x), Standard workshop (1x/yr), Premium custom training

Notes

- Technology choices (e.g., triplestore engine, OCR stack) may vary by deployment; equivalents are supported if they meet performance/compliance targets.

- Premium AI modules are **assistive**; final curation is human-led with logged approvals.

- All figures are defaults and can be tuned in the Statement of Work.