--- INDEX.md ---

BlueSphere — Product Requirements Document Suite

Author: Mark Lindon
Version: v0.35.0
Last Updated: 2025-09-06

This PRD suite expands the original single-file PRD into a set of focused documents. Start here.

Contents

> Tip: If you prefer a single file, open **_COMPILED_PRD.md** in this folder.

--- 01_Vision_Principles.md ---

01 — Vision & Principles

Purpose. BlueSphere turns open ocean observations into **actionable insight** for a **brighter future**—from heatwaves to currents, from the classroom to policymakers.

Vision

- **Clarity**: Make complex ocean signals (SST, currents) explorable by anyone.
- **Action**: Link insights to decisions: conservation, adaptation, education.
- **Openness**: Open data, open APIs, open docs, collaborative ecosystem.

Environmental Outcomes

- Early signals for **marine heatwaves** & **coral bleaching** risk.
- Awareness of **current patterns** affecting fisheries & storms.
- Scalable **education** and **citizen science** participation.

Product Principles

1. **Truthful & transparent**: clear metadata, provenance, QC flags.
2. **Accessible by default**: A11y, mobile-first, low-bandwidth options.
3. **Explain, not just show**: annotations, story panes, glossary.
4. **Responsible**: data ethics, uncertainty, non-alarmist framing.
5. **Composable**: APIs + tiles for downstream reuse.

--- 02_Personas_UseCases.md ---

02 — Personas & Use Cases

Personas

- **Marine Scientist**: detects anomalies; validates models; exports data.
- **Coastal Planner / Policy Maker**: screens risk areas; downloads reports.
- **NGO Program Manager**: monitors MPAs; shares public story maps.
- **Educator (K-12 / Univ.)**: classroom modules; quizzes; lesson embeds.
- **Student / Citizen Scientist**: asks questions; annotates; shares views.
- **Journalist**: quick insight, credible sourcing, embeddable charts.
- **Developer**: integrates tiles/APIs into other apps.

Key Use Cases (selected)

1. **Heatwave Watch**: Track of SST anomaly > +2°C for 5+ days.
2. **Reef Risk Pulse**: Current + SST layers near reefs with alert thresholds.
3. **Classroom Explorer**: Guided story with quiz checkpoints.
4. **Policy Brief Export**: PDF/PNG snapshots with citations & date stamps.
5. **API Timeseries**: Pull station SST for research (CSV/JSON).
6. **Predictive Lookahead (Phase 2)**: 14-day SST anomaly outlook with confidence bands.

--- 03_Scope_and_Success_Criteria.md ---

03 — Scope & Success Criteria

In Scope (Phase 1)

- NDBC WTMP ingestion; ERSST v5 monthly grids; ERDDAP pulls (selected).
- Global interactive map: SST heatmap tiles + currents vectors overlay.
- APIs: /status, /stations, /obs, /obs/summary, tiles endpoints.
- Daily auto-update via CI; nightly tiles refresh.
- Docs: data dictionary, API reference, tutorials, brand basics.

Out of Scope (Phase 1)

- User accounts & private data; write access; fine-grained RBAC.
- Complex predictive models in production (Phase 2+).
- Full offline support.

Success Criteria (Phase 1)

- 95%+ of targeted stations ingested & queryable.
- Map loads first meaningful paint < 2s on median connection.
- CI refresh completes nightly < 60 min.
- A11y audit passes WCAG 2.1 AA critical checks.

--- 04_Functional_Requirements.md ---

04 — Functional Requirements

Ingestion

- Fetch NDBC realtime2 text; parse WTMP; handle `MM` missing; UTC stamps.
- ERSST v5 NetCDF discovery; per-month import; anomalies computed or sourced.
- ERDDAP adapter: tabledap/griddap pulls with retries & backoff.
- Job tracking (`JobRun`): id, source, started, ended, status, metrics.

Storage

- Postgres schemas: `stations`, `buoy_obs`, `job_runs` (see 05_Data_Specs).
- Indexing on (`station_id`, `time`), bbox indexes for lat/lon.

API

- `/status` — health endpoints
- `/stations` — station metadata & extents
- `/obs` — observation query (bbox, station, time, paging)
- `/obs/summary` — aggregations by day (count/avg/min/max)
- Tiles: `/tiles/sst/{z}/{x}/{y}.png`, `/tiles/currents/{z}/{x}/{y}.mvt`

Frontend

- Map view: layer controls (SST, currents, anomalies). Legend + units.
- Time controls: date picker + scrubbing.
- Station panel: metadata + sparkline timeseries.
- Export: PNG snapshot with caption & citation.

Education (Phase 2)

- Story pages with “Explain like I’m 12/18/Pro” toggles.
- Quiz widgets; shareable classroom links.

Chatbot (Phase 2)

- Q&A; over curated knowledge; cites sources; avoids speculation.

Predictive (Phase 2)

- 7–14 day outlook; uncertainty bands; model card (see 13_Predictive...).

--- 05_Data_Specs_and_Schemas.md ---

05 — Data Specs & Schemas

Station Schema (`stations`)

- `station_id` (text, pk) — NDBC code or synthetic id
- `name` (text)
- `lat` (float), `lon` (float)
- `provider` (enum: NDBC, ERDDAP, ERSST, ...)
- `first_obs` (timestamptz), `last_obs` (timestamptz)

Observation Schema (`buoy_obs`)

- `id` (bigserial, pk)
- `station_id` (fk stations.station_id)
- `time` (timestamptz, indexed)
- `sst_c` (float)
- `qc_flag` (int, 0=ok, >0 vendor-specific)
- `lat` (float), `lon` (float)
- `source` (text)

Indexes: `(station_id, time)`, `gist(lat, lon)` or btree on ranges.

JobRun (`job_runs`)

- `id`, `source`, `started`, `ended`, `status` (ok/failed), `rows_ingested`, `error`

Data Quality & QC

- Treat `MM` and sentinel values as nulls.
- Drop obviously bad SST (e.g., < -3°C or > 40°C) unless flagged for analysis.
- Track ingest counts vs vendor counts; alert on deltas > threshold.

Retention & Lineage

- Raw files cached 90 days; derived tiles rebuild on schedule.
- Provenance recorded in `job_runs` + artifact hashes.
- Metadata includes units, CRS (EPSG:3857 for tiles; WGS84 for points).

Units & Conversions

- Temperatures in °C; display toggle °C/°F in UI (Phase 2).

--- 06_API_Specification.md ---

06 — API Specification

Base URL (dev): `http://localhost:8000`

GET /stations

- **Query**: `bbox=minLon,minLat,maxLon,maxLat` (optional)

- **Returns**: array of stations
```json
[{"station_id":"41001","name":"East Hatteras","lat":34.7,"lon":-72.7,"provider":"NDBC"}]
```

GET /obs

- **Query**: `bbox`, `start`, `end`, `station`, `limit` (<=5000), `offset`
- **Returns**: observations ordered by time desc
```json
[{"station_id":"41001","time":"2025-09-06T00:00:00Z","sst_c":29.1,"qc_flag":0,"lat":34.7,"lon":-72.7,"source":"NDBC"}]
```

GET /obs/summary

- **Query**: same filters; groups by day
```json
[{"day":"2025-09-01T00:00:00+00:00","count":144,"avg_sst_c":28.7,"min_sst_c":27.9,"max_sst_c":29.5}]
```

Tiles

- **SST raster**: `/tiles/sst/{z}/{x}/{y}.png` (Web Mercator)
- **Currents vector**: `/tiles/currents/{z}/{x}/{y}.mvt`

Errors
- 400 invalid params; 404 not found; 500 internal error (with request id).

Rate Limits (Phase 2)
- IP-based quotas; `X-RateLimit-*` headers.

--- 07_UI_UX_Spec.md ---

07 — UI / UX Spec

Information Architecture

- Home → Map → Station panel → Sources → About → Docs
- Secondary: Consistency, Roadmap, Brand, Education

Map Interactions

- Pan/zoom; mouse/touch; scroll zoom.
- Layer control: toggle SST / Anomaly / Currents.
- Legend with units and colorbar (see Appendix B).
- Time control: date picker + scrub bar; live “Now” button.
- Click station → side panel with metadata + sparkline chart + link to API.

States & Empty Cases

- Loading spinner; “No data for this date/area” graceful message.
- Offline banner if API unreachable.

Accessibility

- Keyboard nav: focus rings on buttons/links; tab order logical.
- ARIA labels for controls; high-contrast mode.
- Dark mode with automatic and manual toggle.

Export

- “Share view” copies URL + captures PNG of map with legend and timestamp.

--- 08_Accessibility_I18N.md ---

08 — Accessibility & Internationalization

WCAG Targets

- Level AA for color contrast and keyboard operability.
- Descriptive alt text for charts where feasible (data summaries).

Keyboard Shortcuts (Phase 2)

- `f` focus search, `l` toggle legend, `t` open time controls.

Localization

- Default: English.
- Phase 2: String catalog with i18n JSON; RTL-ready layout.

--- 09_NonFunctional_Requirements.md ---

09 — Non-Functional Requirements

Performance

- TTI < 3s on median network for first map load.
- Tiles served via CDN; cache-control: max-age=86400.
- API P95 < 300ms for cached queries; < 600ms for cold reads.

Availability

- SLO: 99.9% uptime; error budget 43m/month.

Security

- HTTPS everywhere (prod); CORS allow-list.
- Input validation; parameterized SQL; no secrets in client.
- Dependency scanning (Snyk/GH Dependabot).

Privacy

- No PII; public datasets; logs scrub IPs (or retain short TTL).

Observability

- Metrics: request rate, latency, error rate, tile cache hit ratio.
- Traces for API handlers; structured JSON logs with request id.

--- 10_Testing_and_Quality.md ---

10 — Testing & Quality

Test Pyramid

- **Unit**: parsers, utils, schema validators.
- **Integration**: ingestion → DB → API.
- **E2E (Playwright)**: map renders, tiles respond, API healthy.

Data Validation

- Range checks on SST; missing value handling; station existence.
- Ingest vs vendor counts within ±5% threshold → else alert.

Performance & Load

- Tiles QPS tests; API soak for 24h.
- Lighthouse CI: performance ≥ 90, accessibility ≥ 95.

Release Gates

- All tests green; CHANGELOG updated; docs build passes.

--- 11_Analytics_Metrics.md ---

11 — Analytics & Metrics

Product Metrics

- DAU/WAU; session duration; map interactions per session.
- Layer toggles; region interest; export usage.
- Education mode completions; quiz scores (anonymized).

Data Metrics

- Stations coverage; ingest recency; job success rate; tile build time.

Environmental Impact Proxies

- Heatwave alerts generated; NGO & educator adoption; citations.

--- 12_Operations_SRE_Runbooks.md ---

12 — Operations, SRE & Runbooks

Runbooks

- Ingestion failure: re-run job with `--since` and inspect `job_runs`.
- Tile cache cold: trigger rebuild workflow; verify CDN headers.
- API degradation: roll back to last green; check error rate & DB health.

On-call

- Pager policy for API 5xx > threshold; tiles error spike.

Backups

- Nightly Postgres snapshots; 7/30/180 day retention tiers.

Config Management

- Env via secrets store; config-as-code; versioned workflows.

--- 13_Predictive_Modeling_Spec.md ---

13 — Predictive Modeling Spec (Phase 2)

Objectives

- 7–14 day SST anomaly outlook with uncertainty; regional focus options.

Data

- ERSST historical; NDBC station series; optional reanalysis features.

Methods (progressive)

1. **Baselines**: climatology, persistence, simple AR.
2. **Classical**: ARIMA/ETS; Prophet-style trend/seasonality.
3. **ML**: gradient boosting for regional signals.
4. **DL (experimental)**: seq2seq or temporal CNN.

Evaluation

- Metrics: RMSE/MAE, CRPS for probabilistic forecasts.
- Cross-validation by region and season.

Model Card

- Purpose, data, limitations, failure modes, update cadence.

MLOps

- Versioned datasets; reproducible notebooks; drift monitors; canary rollouts.

--- 14_Content_Education_Community.md ---

14 — Content, Education & Community

Content

- Explainers: SST, currents, anomalies, marine heatwaves.
- Stories: regional spotlights; student projects.

Education

- Modules by grade band; teacher guides; CC BY licensing where possible.
- Embed codes; printable one-pagers.

Community

- Code of Conduct; contribution guide; issue templates; Discussions.

--- 15_Branding_and_Narrative.md ---

15 — Branding & Narrative

- Logo usage; dark/light variants; wordmark; favicons.
- Tone: credible, clear, hopeful.
- Visuals: ocean blues, clean typography, accessible color ramps.
- Social previews: OG/Twitter images with date-stamped maps.

--- 16_Risks_Dependencies_Assumptions.md ---

16 — Risks, Dependencies, Assumptions

Risks

- Vendor API changes; dataset outages; contradictory sources.
- Misinterpretation of predictions; reputational risk.
- Cost spikes from tile generation or egress.

Mitigations

- Abstraction layer (ERDDAP adapter); fallbacks; caching.
- Clear disclaimers; model cards; uncertainty visuals.
- Budgets, autoscaling guards, CDN.

Assumptions

- Public data remains accessible under current licenses.
- Education partners can co-create modules.

--- 17_Roadmap_Release_Plan.md ---

17 — Roadmap & Release Plan

Milestones

- **M1 (Month 1-2)**: Ingestion + APIs + Map MVP; CI nightly refresh.
- **M2 (Month 3-4)**: Anomalies, station panels, exports; Sources/About.
- **M3 (Month 5-6)**: Education basics, community pages; beta launch.
- **M4 (Phase 2)**: Predictive models + chatbot; classroom mode.
- **M5 (Phase 3)**: Integrations & API ecosystem; partnerships.

Launch Criteria

- Data coverage ≥ 95% target set, uptime ≥ 99.9%, A11y & perf budgets met.
- Docs complete; tests green; issues triaged.

--- 18_Legal_Privacy_Licensing.md ---

18 — Legal, Privacy, Licensing

- Public domain NOAA data; cite sources; link to terms.
- Privacy: no PII; anonymized analytics with opt-out.
- Terms: not for navigation or life-safety decisions.
- Licenses: repo under MIT/Apache-2; content under CC BY 4.0 where noted.

--- Appendix_A_Glossary.md ---

Appendix A — Glossary

- **SST**: Sea-Surface Temperature.
- **Anomaly**: difference from climatological baseline.
- **ERDDAP**: data server for gridded/tabular datasets.
- **ERSST**: Extended Reconstructed SST (NOAA/NCEI).
- **WTMP**: water temperature variable in NDBC feeds.
- **Web Mercator (EPSG:3857)**: standard web map projection.

--- Appendix_B_Tiling_Projections_Colorbars.md ---

Appendix B — Tiling, Projections & Colorbars

Tiling

- Scheme: XYZ tiles, 256px, Web Mercator.
- Zoom range: z=0..6 (global overview) → extend as needed.
- SST PNG tiles; currents MVT vectors (speed/dir).

Projections

- Data stored in WGS84 (lat/lon); rendered as EPSG:3857.

Colorbars

- Sequential for SST: deep blue → aqua → yellow for warmer waters.

- Diverging for anomalies: blue (negative) ↔ white ↔ red (positive).
- Ensure contrast & colorblind-friendly ramps; include tick marks with units.